PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/24, 15/55	A1	(11) International Publication Number: WO 99/14314 (43) International Publication Date: 25 March 1999 (25.03.99)

(21) International Application Number: PCT/AU98/00743

(22) International Filing Date: 11 September 1998 (11.09.98)

(30) Priority Data:
PO 9108 12 September 1997 (12.09.97) AU
PP 2509 20 March 1998 (20.03.98) AU

(71) Applicants *(for all designated States except US)*: COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION [AU/AU]; Limestone Avenue, Campbell, ACT 2612 (AU). THE AUSTRALIAN NATIONAL UNIVERSITY [AU/AU]; Acton, ACT 2601 (AU). GOODMAN FIELDER LIMITED [AU/AU]; Level 42, Grosvenor Place, Sydney, NSW 2000 (AU). GROUPE LIMAGRAIN PACIFIC PTY. LIMITED [AU/AU]; Level 31, 1 O'Connell Street, Sydney, NSW 2000 (AU).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: LI, Zhongyi [CN/AU]; 63 Campaspe Circuit, Kaleen, ACT 2617 (AU). MORELL, Matthew [AU/AU]; 33 Wangara Street, Aranda, ACT 2614 (AU). RAHMAN, Sadequr [AU/AU]; 46 Scarlett Street, Melba, ACT 2615 (AU).

(74) Agent: GRIFFITH HACK; 509 St. Kilda Road, Melbourne, VIC 3004 (AU).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: REGULATION OF GENE EXPRESSION IN PLANTS

(57) Abstract

The present invention relates to a nucleic acid sequence encoding an enzyme of the starch biosynthetic pathway in a cereal plant, wherein the enzyme is selected from the group consisting of starch branching enzyme I, starch branching enzyme II, starch soluble synthase I, and debranching enzyme, with the proviso that the enzyme is not soluble starch synthase I of rice, or starch branching enzyme I of rice or maize.

wheat bp
Seq. ID Oligos
5 2687 BE I cDNA
7 319 BE I 3' untranslated
8 4890 BE I + promoter
9 6228 BE I gene
10 11463 BE II gene
11 2662 SS I cDNA
12

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

4/PRTS

09/674817

PCT/AU98/00743

529 Rec'd PCT/PTC 06 NOV 2000

REGULATION OF GENE EXPRESSION IN PLANTS

This invention relates to methods of modulating the expression of desired genes in plants, and to DNA sequences and genetic constructs for use in these methods. In particular, the invention relates to methods and constructs for targeting of expression specifically to the endosperm of the seeds of cereal plants such as wheat, and for modulating the time of expression in the target tissue. This is achieved by the use of promoter sequences from enzymes of the starch biosynthetic pathway. In a preferred embodiment of the invention, the sequences and/or promoters are those of starch branching enzyme I, starch branching enzyme II, soluble starch synthase I, and starch debranching enzyme, all derived from *Triticum tauschii*, the D genome donor of hexaploid bread wheat.

A further preferred embodiment relates to a method of identifying variations in the characteristics of plants.

BACKGROUND OF THE INVENTION

Starch is an important constituent of cereal grains and of flours, accounting for about 65-67% of the weight of the grain at maturity. It is produced in the amyloplast of the grain endosperm by the concerted action of a number of enzymes, including ADP-Glucose pyrophosphorylase (EC 2.7.7.27), starch synthases (EC 2.4.1.21), branching enzymes (EC 2.4.1.18) and debranching enzymes (EC 3.2.1.41 and EC 3.2.1.68) (Ball *et al*, 1996; Martin and Smith, 1995; Morell *et al*, 1995). Some of the proteins involved in the synthesis of starch can be recovered from the starch granule (Denyer *et al*, 1995; Rahman *et al*, 1995).

Most wheat cultivars normally produce starch containing 25% amylose and 75% amylopectin. Amylose is composed of large linear chains of α (1-4) linked α-D-glucopyranosyl residues, whereas amylopectin is a branching form of α-glycan linked by α (1-6) linkages. The ratio of amylose and amylopectin, the branch chain length and the

number of branch chains of amylopectin are the major factors which determine the properties of wheat starch.

Starch with various properties has been widely used in industry, food science and medical science. High amylose wheat can be used for plastic substitutes and in paper manufacture to protect the environment; in health foods to reduce bowel cancer and heart disease; and in sports foods to improve the athletes' performance. High amylopectin wheat may be suitable for Japanese noodles, and is used as a thickener in the food industry.

Wheat contains three sets of chromosomes (A, B and D) in its very large genome of about 10^{10} base pairs (bp). The donor of the D genome to wheat is *Triticum tauschii*, and by using a suitable accession of this species the genes from the D genome can be studied separately (Lagudah *et al*, 1991).

There is comparatively little variation in starch structure found in wheat varieties, because the hexaploid nature of wheat prevents mutations from being readily identified. Dramatic alterations in starch structure are expected to require the combination of homozygous recessive alleles from each of the 3 wheat genomes, A, B and D. This requirement renders the probability of finding such mutants in natural or mutagenised populations of wheat very low. Variation in wheat starch is desirable in order to enable better tailoring of wheat starches for processing and end-user requirements.

Key commercial targets for the manipulation of starch biosynthesis are:

1. "Waxy" wheats in which amylose content is decreased to insignificant levels. This outcome is expected to be obtained by eliminating granule-bound starch synthase activity.

2. High amylose wheats, expected to be obtained by suppressing starch branching enzyme-II activity.

3. Wheats which continue to synthesise starch at elevated temperatures, expected to be obtained by

identifying or introducing a gene encoding a heat-stable soluble starch synthase.

4. "Sugary types" of wheat which contain increased amylose content and free sugars, expected to be obtained by manipulating an isoamylase-type debranching enzyme.

There are two general strategies which may be used to obtain wheats with altered starch structure:

(a) using genetic engineering strategies to suppress the activity of a specific gene, or to introduce a novel gene into a wheat line; and

(b) selecting among existing variation in wheat for missing ("null") or altered alleles of a gene in each of the genomes of wheat, and combining these by plant breeding.

However, in view of the complexity of the gene families, particularly starch branching enzyme I (SBE I), without the ability to target regions which are unique to genes expressed in endosperm, modification of wheat by combination of null alleles of several enzymes in general represents an almost impossible task.

Branching enzymes are involved in the production of glucose α-1,6 branches. Of the two main constituents of starch, amylose is essentially linear, but amylopectin is highly branched; thus branching enzymes are thought to be directly involved in the synthesis of amylopectin but not amylose. There are two types of branching enzymes in plants ,starch branching enzyme I (SBE I) and starch branching enzyme II (SBE II), and both are about 85 kDa in size. At the nucleic acid level there is about 65% sequence identity between types I and II in the central portion of the molecules; the sequence identity between SBE I from different cereals is about 85% overall (Burton *et al*, 1995; Morell *et al*, 1995).

In cereals, SBE I genes have so far been reported only for rice (Kawasaki *et al*, 1991; Rahman *et al*, 1997). A cDNA sequence for wheat SBE I is available on the GenBank

database (Accession No. Y12320; Repellin A., Nair R.B., Baga M., and Chibbar R.N.: Plant Gene Register PGR97-094, 1997). As far as we are aware, no promoter sequence for wheat SBE I has been reported.

We have characterised an SBE I gene, designated *wSBE I-D2*, from *Triticum tauschii*, the donor of the D genome to wheat (Rahman *et al*, 1997). This gene encoded a protein sequence which had a deletion of approximately 65 amino acids at the C-terminal end, and appeared not to contain some of the conserved amino acid motifs characteristic of this class of enzyme (Svensson, 1994). Although *wSBE I-D2* was expressed as mRNA, no corresponding protein has yet been found in our analysis of SBE I isoforms from the endosperm, and thus it is possible that this gene is a transcribed pseudogene.

Genes for SBE II are less well characterised; no genomic sequences are available, although SBE II cDNAs from rice (Mizuno *et al*, 1993; Accession No. D16201) and maize (Fisher *et al*, 1993; Accession No. L08065) have been reported. In addition, a cDNA sequence for SBE II from wheat is available on the GenBank database (Nair *et al*, 1997; Accession No. Y11282); although the sequences are very similar to those reported herein, there are differences near the N-terminal of the protein, which specifies its intracellular location. No promoter sequences have been reported, as far as we are aware.

Wheat granule-bound starch synthase (GBSS) is responsible for amylose synthesis, while wheat branching enzymes together with soluble starch synthases are considered to be directly involved in amylopectin biosynthesis. A number of isoforms of soluble and granule-bound starch synthases have been identified in developing wheat endosperm (Denyer *et al*, 1995). There are three distinct isoforms of starch synthases, 60 kDa, 75-77 kDa and 100-105 kDa, which exist in the starch granules (Denyer *et al*, 1995; Rahman *et al*, 1995). The 60 kDa GBSS is the product of the wx gene. The 75-77 kDa protein is a wheat

soluble starch synthase I (SSSI) which is present in both the soluble fraction and the starch granule-bound fraction of the endosperm. However, the 100-105 kDa proteins, which are another type of soluble starch synthase, are located only in starch granules (Denyer *et al*, 1995; Rahman *et al*, 1995). To our knowledge there has been no report of any complete wheat SSS I sequence, either at the protein or the nucleotide level.

Both cDNA and genomic DNA encoding a soluble starch synthase I of rice have been cloned and analysed (Baba *et al*, 1993; Tanaka *et al*, 1995). The cDNAs encoding potato soluble starch synthase SSSII and SSSIII and pea soluble starch synthase SSSII have also been reported (Edwards *et al*, 1995; Marshall *et al*, 1996; Dry *et al*, 1992). However, corresponding full length cDNA sequences for wheat have hitherto not been available, although a partial cDNA sequence (Accession No. U48227) has been released to the GenBank database.

Approach (b) referred to above has been demonstrated for the gene for granule-bound starch synthase. Null alleles on chromosomes 7A, 7D and 4A were identified by the analysis of GBSS protein bands by electrophoresis, and combined by plant breeding to produce a wheat line containing no GBSS, and no amylose (Nakamura *et al*, 1995). Subsequently, PCR-based DNA markers have been identified, which also identify null alleles for the GBSS loci on each of the three wheat genomes. Despite the availability of a considerable amount of information in the prior art, major problems remain. Firstly, the presence of three separate sets of chromosomes in wheat makes genetic analysis in this species extraordinarily complex. This is further complicated by the fact that a number of enzymes are involved in starch synthesis, and each of these enzymes is itself present in a number of forms, and in a number of locations within the plant cell. Little, if any, information has been available as to which specific form of each enzyme is expressed in endosperm. For wheat, a limited

amount of nucleic acid sequence information is available, but this is only cDNA sequence; no genomic sequence, and consequently no information regarding promoters and other control sequences, is available. Without being able to demonstrate that the endosperm-specific gene within a family has been isolated, such sequence information is of limited practical usefulness.

SUMMARY OF THE INVENTION

In this application we report the isolation and identification of novel genes from *T. tauschii*, the D-genome donor of wheat, that encode SBE I, SBE II, a 75 kDa SSS I, and an isoamylase-type debranching enzyme (DBE). Because of the very close relationship between *T. tauschii* and wheat, as discussed above, results obtained with *T. tauschii* can be directly applied to wheat with little if any modification. Such modification as may be required represents routine trial and error experimentation. Sequences from these genes can be used as probes to identify null or altered alleles in wheat, which can then be used in plant breeding programmes to provide modifications of starch characteristics. The novel sequences of the invention can be used in genetic engineering strategies or to introduce a desired gene into a host plant, to provide antisense sequences for suppression of one or more specific genes in a host plant, in order to modify the characteristics of starch produced by the plant.

By using *T. tauschii*, we have been able to examine a single genome, rather than three as in wheat, and to identify and isolate the forms of the starch synthesis genes which are expressed in endosperm. By addressing genomic sequences we have been able to isolate tissue-specific promoters for the relevant genes, which provides a mechanism for simultaneous manipulation of a number of genes in the endosperm. Because *T. tauschii* is so closely related to wheat, results obtained with this model system are directly applicable to wheat, and we have confirmed this experimentally. The genomic sequences which we have

determined can also be used as probes for the identification and isolation of corresponding sequences, including promoter sequences, from other cereal plant species.

In its most general aspect, the invention provides a nucleic acid sequence encoding an enzyme of the starch biosynthetic pathway in a cereal plant, said enzyme being selected from the group consisting of starch branching enzyme I, starch branching enzyme II, starch soluble synthase I, and debranching enzyme, with the proviso that the enzyme is not soluble starch synthase I of rice, or starch branching enzyme I of rice or maize.

Preferably the nucleic acid sequence is a DNA sequence, and may be genomic DNA or cDNA. Preferably the sequence is one which is functional in wheat. More preferably the sequence is derived from a *Triticum* species, most preferably *Triticum tauschii*.

Where the sequence encodes soluble starch synthase, preferably the sequence encodes the 75 kD soluble starch synthase of wheat.

Biologically-active untranslated control sequences of genomic DNA are also within the scope of the invention. Thus the invention also provides the promoter of an enzyme as defined above.

In a preferred embodiment of this aspect of the invention, there is provided a nucleic acid construct comprising a nucleic acid sequence of the invention, a biologically-active fragment thereof, or a fragment thereof encoding a biologically-active fragment of an enzyme as defined above, operably linked to one or more nucleic acid sequences facilitating expression of said enzyme in a plant, preferably a cereal plant. The construct may be a plasmid or a vector, preferably one suitable for use in the transformation of a plant. A particularly suitable vector is a bacterium of the genus *Agrobacterium*, preferably *Agrobacterium tumefaciens*. Methods of transforming cereal plants using *Agrobacterium tumefaciens* are known; see for example Australian Patent No. 667939 by Japan Tobacco Inc.,

International Patent Application Number PCT/US97/10621 by Monsanto Company and Tingay *et al* (1997).

In a second aspect, the invention provides a nucleic acid construct for targeting of a desired gene to endosperm of a cereal plant, and/or for modulating the time of expression of a desired gene in endosperm of a cereal plant, comprising one or more promoter sequences selected from SBE I promoter, SBE II promoter, SSS I promoter, and DBE promoter, operatively linked to a nucleic acid sequence encoding a desired protein, and optionally also operatively linked to one or more additional targeting sequences and/or one or more 3' untranslated sequences.

The nucleic acid encoding the desired protein may be in either the sense orientation or in the antisense orientation. Preferably the desired protein is an enzyme of the starch biosynthetic pathway. For example, the antisense sequences of GBSS, starch debranching enzyme, SBE II, low molecular weight glutenin, or grain softness protein I, may be used. Preferred sequences for use in sense orientation include those of bacterial isoamylase, bacterial glycogen synthase, or wheat high molecular weight glutenin Bx17. It is contemplated that any desired protein which is encoded by a gene which is capable of being expressed in the endosperm of a cereal plant is suitable for use in the invention.

In a third aspect, the invention provides a method of modifying the characteristics of starch produced by a plant, comprising the step of:

(a) introducing a gene encoding a desired enzyme of the starch biosynthetic pathway into a host plant, and/or

(b) introducing an anti-sense nucleic acid sequence directed to a gene encoding an enzyme of the starch biosynthetic pathway into a host plant,

wherein said enzymes are as defined above.

Where both steps (a) and (b) are used, the enzymes in the two steps are different.

Preferably the plant is a cereal plant, more preferably wheat or barley.

As is well known in the art, anti-sense sequences can be used to suppress expression of the protein to which the anti-sense sequence is complementary. It will be evident to the person skilled in the art that different combinations of sense and anti-sense sequences may be chosen so as to effect a variety of different modifications of the characteristics of the starch produced by the plant.

In a fourth aspect, the invention provides a method of targeting expression of a desired gene to the endosperm of a cereal plant, comprising the step of transforming the plant with a construct according to the invention.

According to a fifth aspect, the invention provides a method of modulating the time of expression of a desired gene in endosperm of a cereal plant, comprising the step of transforming the plant with a construct according to the second aspect of the invention.

Where expression at an early stage following anthesis is desired, the construct preferably comprises the SBE II, SSS I or DBE promoters. Where expression at a later stage following anthesis is desired, the construct preferably comprises the SBE I promoter.

While the invention is described in detail in relation to wheat, it will be clearly understood that it is also applicable to other cereal plants of the family Gramineae, such as maize, barley and rice.

Methods for transformation of monocotyledonous plants such as wheat, maize, barley and rice and for regeneration of plants from protoplasts or immature plant embryos are well known in the art. See for example Lazzeri *et al*, 1991; Jahne *et al*, 1991 and Wan and Lemaux, 1994 for barley; Wirtzens *et al*, 1997; Tingay *et al*, 1997; Canadian Patent Application No. 2092588 by Nehra; Australian Patent Application No. 61781/94 by National Research Council of Canada, Australian Patent No. 667939 by Japan Tobacco Co, and International Patent Application Number PCT/US97/10621 by Monsanto Company.

The sequences of ADP glucose pyrophosphorylase from barley (Australian Patent Application No. 65392/94), starch debranching enzyme and its promoter from rice (Japanese Patent Publication No. Kokai 6261787 and Japanese Patent Publication No. Kokai 5317057), and starch debranching enzyme from spinach and potato (Australian Patent Application No. 44333/96) are all known.

Detailed Description of the Drawings

The invention will be described in detail by reference only to the following non-limiting examples and to the figures.

Figure 1 shows the hybridisation of genomic clones isolated from *T. tauschii*.

DNA was extracted from the different clones, digested with *BamHI* and hybridised with the 5′ end of the maize SBE I cDNA. Lanes 1, 2, 3 and 4 correspond to DNA from clones λE1, λE2, λE6 and λE7 respectively. Note that clones λE1 and λE2 give identical patterns, the SBE I gene in λE6 is a truncated form of that in λE1, and λE7 gives a clearly different pattern.

Figure 2 shows the hybridisation of DNA from *T. tauschii*.

DNA from *T. tauschii* was digested with *BamHI* and the hybridisation pattern compared with DNA from λE1 and λE7 digested with the same enzyme. Fragment E1.1 (see Figure 3) from λE1 was used as the probe; it contains some sequences that are over 80% identical to sequences in E7.8. Approximately 25 µg of *T. tauschii* DNA was electrophoresed in lane 1, and 200 pg each of λE1 and λE7 in lanes 2 and 3, respectively.

Figure 3 shows the restriction maps of clone λE1 and λE7. The fragments obtained with EcoRI and BamHI are indicated. The fragments sequenced from λE1 are E1.1, E1.2, a part of E1.7 and a part of E1.5.

Figure 4 shows the comparison of deduced amino acid sequence of wSBE I-D4 cDNA with the deduced amino acid

sequence of rice SBE I (RSBE I; Nakamura *et al*, 1992), maize SBE I (MSBE I; Baba *et al*, 1991), wSBE I-D2 type cDNA (D2 cDNA; Rahman *et al*, 1997), pea SBE II (PESBE II, homologous to maize SBE I; Burton *et al*, 1995), and potato SBE I (POSBE; Cangiano *et al*, 1993). The deduced amino acid sequence of the wSBE I-D4 cDNA is denoted by "D4cDNA". Residues present in at least three of the sequences are identified in the consensus sequence in capitals.

Figure 5 shows the intron-exon structure of wSBE I-D4 compared to the corresponding structures of rice SBE I (Kawasaki *et al*, 1993) and wSBE I-D2 (Rahman *et al*, 1997). The intron-exon structure of wSBE I-D4 is deduced by comparison with the SBE I cDNA reported by Repellin *et al* (1997).

The dark rectangles correspond to exons and the light rectangles correspond to introns. The bars above the structures indicate the percentage identity in sequence between the indicated exons and introns of the relevant genes. Note that intron 2 shares no significant sequence identity and is not indicated.

Figure 6 shows the nucleotide sequence of part of wSBE I-D4, the amino acid sequence deduced from this nucleotide sequence, and the N-terminal amino acid sequence of the SBE I purified from the wheat endosperm (Morell *et al*, 1997).

Figure 7 shows the hybridisation of SBE I genomic clones with the following probes,

A. wSBE I-D45 (derived from the 5' end of the gene and including sequence from fragments E1.1 and E1.7), and

B. wSBE I-D43 (derived from the 3' end of the gene and containing sequences from fragment E1.5). For panel A, the tracks 1-13 correspond to clones λE1, λE2, λE6, λE7, λE9, λE14, λE22, λE27, Molecular weight markers, λE29, λE30, λE31 and λE52. For panel B, tracks 1-12 correspond to clones λE1, λE2, λE6, λE7, λE9, λE14, λE22, λE27, λE29, λE30, λE31 and λE52. Note that clones λE7 and λE22 do not

hybridise to either of the probes and are wSBE I-D2 type genes. Also note that clone λE30 contains a sequence unrelated to SBE I. The size of the molecular weight markers in kb is indicated. Clones λE7 and λE22 do hybridise with a probe from E1.1. which is highly conserved between wSBE I-D2 and wSBE I-D4.

Figure 8 shows the alignment of cDNA clones to obtain the sequence represented by wSBE I-D4 cDNA. BED4 and BED5 were obtained from screening the cDNA library with maize BEI (Baba *et al*, 1991). BED1, 2 and 3 were obtained by RT-PCR using defined primers.

Figure 9a shows the expression of Soluble Starch Synthase I (SSS), Starch Branching Enzyme I (BE I) and Starch Branching Enzyme II (BE II) mRNAs during endosperm development.

RNA was purified from leaves, florets prior to anthesis, and endosperm of wheat cultivar Rosella grown in a glasshouse, collected 5 to 8 days after anthesis, 10 to 15 days after anthesis and 18 to 22 days after anthesis, and from the endosperm of wheat cultivar Rosella grown in the field and collected 12, 15 and 18 days after anthesis respectively. Equivalent amounts of RNA were electrophoresed in each lane. The probes were from the coding region of the SM2 SSS I cDNA (from nucleotide 1615 to 1919 of the SM2 cDNA sequence); wSBE I-D43C (see Table I), which corresponds to the untranslated 3' end of wSBE I-D4 cDNA (E1 (3'; and the 5' region of SBE9 (SBE9 (5'), corresponding to the region between nucleotides 743 to 1004 of Genbank sequence Y11282. No hybridisation to RNA extracted from leaves or preanthesis florets was detected.

Figure 9b shows the hybridisation of RNA from the endosperm of the hexaploid *T. aestivum* cultivar "Gabo" with the starch branching enzyme I gene. The probe, wSBEI-D43, is defined in Table 1.

Figure 9c shows the hybridisation of RNA from the endosperm of the hexaploid *T. aestivum* cultivar "Wyuna" with

the starch branching enzyme II gene. The probe, wSBE II-D13, is defined in Table 2.

Figure 9d shows the hybridisation of RNA from the endosperm of the hexaploid *T. aestivum* cultivar "Gabo" with the SSS I gene. The probe spanned the region from nucleotides 2025 to 2497 of the SM2 cDNA sequence shown in SEQ ID No:11.

Figure 9e shows the hybridisation of RNA from the endosperm of the hexaploid *T. aestivum* cultivar "Gabo" with the DBE I gene. The probe, a DBE3' 3'PCR fragment, extends from nucleotide position 281 to 1072 of the cDNA sequence in SEQ ID No:16.

Figure 9f shows the hybridisation of RNA from the endosperm of the hexaploid *T. aestivum* cultivar "Gabo" with the wheat actin gene. The probe was a wheat actin DNA sequence generated by PCR from wheat endosperm cDNA using primers to conserved plant actin sequences.

Figure 9g shows the hybridisation of RNA from the endosperm of the hexaploid *T. aestivum* cultivar "Gabo" with a probe containing wheat ribosomal RNA 26S and 18S fragments (plasmid pta250.2 from Dr Bryan Clarke, CSIRO Plant Industry).

Figure 9h shows the hybridisation of RNA from the hexaploid wheat cultivar "Gabo" with the DBE I probe described in Figure 9e. Lane 1; leaf RNA; lane 2, pre-anthesis floret RNA; lane 3, RNA from endosperm harvested 12 days after anthesis.

Figure 10 shows the comparison of wSBE I-D4 (sr 427.res ck: 6,362,1 to 11,099) and rice SBE I genomic sequence (d10838.em_pl ck: 3,071,1 to 11,700)(Kawasaki *et al*, 1993; Accession Number D10838) using the programs Compares and DotPlot (Devereaux *et al*, 1984). The programs used a window of 21 bases with a stringency of 14 to register a dot.

Figure 11 shows the hybridisation of wheat DNA from chromosome-engineered lines using the following probes:

A. wSBE I-D45 (from the 5' end of the gene),

B. wSBE I-D43 (from the 3' end of the gene), and

C. wSBE I-D4R (repetitive sequence approximately 600 bp 3' to the end of wSBE I-D4 sequence.

N7AT7B, no 7A chromosome, four copies of 7B chromosome; N7BT7D, no 7B chromosome, four copies of 7D chromosome; NTDT7A, no 7D chromosome, four copies of 7A chromosome. The chromosomal origin of hybridising bands is indicated.

Figure 12 shows the hybridisation of genomic clones F1, F2, F3 and F4 with the entire SBE-9 sequence. The DNA from the clones was purified and digested with either *Bam*HI or *Eco*RI, separated on agarose, blotted onto nitrocellulose and hybridised with labelled SBE-9 (a SBE II type cDNA). The pattern of hybridising bands is different in the four isolates.

Figure 13a shows the N-terminal sequence of purified SBE II from wheat endosperm as in Morell *et al*, (1997).

Figure 13b shows the deduced amino acid sequence from part of wSBE II-D1 that encodes the N-terminal sequence as described in Morell *et al*, (1997).

Figure 14 shows the deduced exon-intron structure for a part of wSBE II-D1. The scale is marked in bases. The dark rectangles are exons.

Figure 15 shows the hybridisation of DNA from chromosome engineered lines of wheat (cultivar Chinese Spring) with a probe from nucleotides 550-850 from SBE-9. The band of approximately 2.2 kb is missing in the line in which chromosome 2D is absent.

T2BN2A: four copies of chromosome 2B, no copies of chromosome 2A;

T2AN2B: four copies of chromosome 2A, no copies of chromosome 2B;

T2AN2D: four copies of chromosome 2A, no copies of chromosome 2D.

Figure 16 shows the N-terminal sequence of SSS I protein isolated from starch granules (Rahman *et al*, 1995) and deduced amino acid sequence of part of Sm2.

Figure 17 shows the hybridisation of genomic clones sg1, 3, 4, 6 and 11 with the cDNA clone (sm2) for SSS I. DNA was purified from indicated genomic clones, digested with *Bam*HI or *Sac*I and hybridised to sm2. Note that the hybridisation patterns for sg1, 3 and 4 are clearly different from each other.

Figure 18 shows a comparison of the intron/exon structures of the wheat and rice soluble starch synthase genomic sequences. The dark rectangles indicate exons and the light rectangles represent introns.

Figure 19 shows the hybridisation of DNA from chromosome engineered lines of wheat (cultivar Chinese Spring) digested with *Pvu*II, with the sm2 probe.

N7AT7B: no 7A chromosome, four copies of 7B chromosome;

N7BT7D: no 7B chromosome, four copies of 7D chromosome;

N7DT7A: no 7D chromosome, four copies of 7A chromosome.

A band is missing in the N7BT7A line.

Figure 20a shows the DNA sequence of a portion of the wheat debranching enzyme (WDBE-1) PCR product. The PCR product was generated from wheat genomic DNA (cultivar Rosella) using primers based on sequences conserved in debranching enzymes from maize and rice.

Figure 20b shows a comparison of the nucleotide sequence of wheat debranching enzyme I (WDBE-I) PCR fragment (WHEAT.DNA) with the maize *Sugary-1* sequence (SUGARY.DNA).

Figure 20c shows a comparison between the intron/exon structures of wheat debranching enzyme gene and the maize sugary-1 debranching enzyme gene.

Figure 21a shows the results of Southern blotting of *T. tauschii* DNA with wheat DBE-I PCR product. DNA from *T. tauschii* was digested with *Bam*HI , electrophoresed,

blotted and hybridised to the wheat DBE-I PCR product described in Figure 20a. A band of approximately 2 kb hybridised.

Figure 21b shows Chinese Spring nullisomic/ tetrasomic lines probed with probes from the DBE gene. Panel (I) shows hybridisation with a fragment spanning the region from nucleotide 270 to 465 of the cDNA sequence shown in SEQ ID No:16 from the central region of the DBE gene. Panel (II) shows hybridisation with a probe from the 3′ region of the gene, from nucleotide 281 to 1072 of the cDNA sequence given in SEQ ID No:16.

Figures 22a to 22e show diagrammatic representations of the DNA vectors used for transient expression analysis. In each of the sequences the N-terminal methionine encoding ATG codon is shown in bold.

Figure 22a shows a DNA construct pwsssIpro1gfpNOT containing a 1042 base pair region of the wheat soluble starch synthase I promoter (wSSSIpro1, from -1042 to -1, SEQ ID No:18) fused to the green fluorescent protein (GFP) reporter gene.

Figure 22b shows a DNA construct pwsssIpro2gfpNOT containing a 3914 base pair region of the wheat soluble starch synthase I promoter (wSSSIpro2, from -3914 to -1, SEQ ID No:18) fused to the green fluorescent protein (GFP) reporter gene.

Figure 22c shows a DNA construct psbeIIpro1gfpNOT containing an 1203 base pair region of the wheat starch branching enzyme II promoter (sbeIIpro1, from 1 to 1023 SEQ ID No:10 fused to the green fluorescent protein (GFP) reporter gene.

Figure 22d shows a DNA construct psbeIIpro2gfpNOT containing a 1353 base pair region of the wheat starch branching enzyme II promoter and transit peptide coding region (sbeIIpro2, regions 1-1203, 1204 to 1336 and 1664 to 1680 of SEQ ID No:10 fused to the green fluorescent protein (GFP) reporter gene.

Figure 22e shows a DNA construct pact_jsgfg_nos

containing the plasmid backbone of pSP72 (Promega), the rice *ActI* actin promoter (McElroy et al. 1991), the GFP gene (Sheen et al. 1995) and the *Agrobacterium tumefaciens* nopaline synthase (nos) terminator (Bevan et al. 1983).

Figure 23 shows T DNA constructs for stable transformation of rice by *Agrobacterium.* The backbone for each plasmid is p35SH-iC (Wang et al 1997). The various promoter-GFP-Nos regions inserted are shown in (a), (b), (c) and (d) respectively, and are described in detail in Example 24. Each of these constructs was inserted into the NotI site of p35SH-iC using the NotI flanking sites at each end of the promoter-GFP-Nos regions. The constructs were named (a) p35SH-iC-BEIIpro1_GFP_Nos, (b) p35SH-iC-BEIIpro2_GFP_Nos (c) p35SH-iC-SSIpro1_GFP_Nos and (d) p35SH-iC-SSIpro2_GFP_Nos

Figure 24 illustrates the design of 15 intron-spanning BE II primer sets. Primers were based on wSBE II-D1 sequence (SEQ ID No:10), and were designed such that intron sequences in the wSBE II-D1 sequence (deduced from Figure 13b and Nair *et al*, 1997; Accession No. Y11282) were amplified by PCR.

Figure 25 shows the results of amplification using the SBE II-Intron 5 primer set (primer set 6: sr913F and WBE2E6 R) on various diploid, tetraploid and hexaploid wheats.

i) *T.boeodicum* (A genome diploid)
ii) *T.tauschii* (D genome diploid)
iii) *T.aestivum* cv. Chinese Spring ditelosomic line 2AS (lacking chromosome arm 2AL)
iv) Crete 10 (AABB tetraploid)
v) *T. aestivum* cv Rosella (hexaploid)

The horizontal axis indicates the size of the product in base pairs, the vertical axis shows arbitrary fluorescence units. The various arrows indicate the products of different genomes: A, A genome, B, B genome, D, D genome, U, unassigned additional product.

Figure 26 shows the results obtained by amplification using the SBE II-Intron 10 primer set (primer set 11: da5.seq and WBE2E11R on the wheat lines:

(i) *T. aestivum* cv. Chinese Spring ditelosomic line 2AS.

(ii) *T. aestivum* Chinese Spring nullisomic/tetrasomic line N2BT2A.

(iii) *T. aestivum* Chinese Spring nullisomic/tetrasomic line N2DT2B.

The horizontal axis indicates the size of the product in base pairs, the vertical axis shows arbitrary fluorescence units. The various arrows indicate the products of different genomes: A, A genome, B, B genome, D, D genome.

Figure 27 shows the results of transient expression assays typical of each promoter and target tissue. The photographs (40 x magnification) of representative tissue resulting from the transient expression assays typical of each promoter and target tissue revealed under a Leica microscope with blue light illumination. Photographs were taken 48 to 72 hours after tissue bombardment. The promoter constructs are listed as follows, (with the panels showing endosperm, embryo and leaf expression listed in respective order): pact_jsgfp_nos (panels a,g and m); pwsssIpro1gfpNOT (panels b, h and n); pwsssIpro2gfpNOT (panels c, i and o); psbeIIpro1gfpNOT (panels d, j and p); psbeIIpro2gfpNOT (panels e, k and q); pZLgfpNOT (Panels f, l and r).

Example 1 Identification of Gene Encoding SBE I

Construction of Genomic Library and Isolation of Clones

The genomic library used in this study was constructed from *Triticum tauschii*, var. strangulata, accession number CPI 100799. Of all the accessions of *T. tauschii* surveyed, the genome of CPI 100799 is the most closely related to the D genome of hexaploid wheat.

Triticum tauschii, var strangulata (CPI accession number 110799) was kindly provided by Dr E Lagudah. Leaves were isolated from plants grown in the glasshouse.

DNA was extracted from leaves of *Triticum tauschii* using published methods (Lagudah *et al*, 1991), partially digested with *Sau*3A, size fractionated and ligated to the arms of lambda GEM 12 (Promega). The ligated products were used to transfect the methylation-tolerant strain PMC 103 (Doherty *et al*. 1992). A total of 2 x 10^6 primary plaques were obtained with an average insert size of about 15 kb. Thus the library contains approximately 6 genomes worth of *T. tauschii* DNA. The library was amplified and stored at 4°C until required.

Positive plaques in the genomic library were selected as those hybridising with the 5' end of a maize starch branching enzyme I cDNA (Baba *et al*, 1991) using moderately stringent conditions as described in Rahman *et al*, (1997).

Preparation of Total RNA from Wheat

Total RNA was isolated from leaves, pre-anthesis pericarp and different developmental stages of wheat endosperm of the cultivar, Hartog and Rosella. This material was collected from both the glasshouse and the field. The method used for RNA isolation was essentially the same as that described by Higgins *et al* (1976). RNA was then quantified by UV absorption and by separation in 1.4% agarose-formaldehyde gels which were then visualized under UV light after staining with ethidium bromide (Sambrook *et al*, 1989).

DNA and RNA analysis

DNA was isolated and analysed using established protocols (Sambrook *et al*, 1989). DNA was extracted from wheat (cv. Chinese Spring) using published methods (Lagudah *et al*, 1991). Southern analysis was performed essentially as described by Jolly *et al* (1996). Briefly, 20 µg wheat

DNA was digested, electrophoresed and transferred to a nylon membrane. Hybridisation was conducted at 42°C in 25% or 50% formamide, 2 x SSC, 6% Dextran Sulphate for 16h and the membrane was washed at 60°C in 2 x SSC for 3 x 1h unless otherwise indicated. Hybridisation was detected by autoradiography using Fuji X-Omat film.

RNA analysis was performed as follows. 10 µg of total RNA was separated in a 1.4% agarose-formaldehyde gel and transferred to a nylon Hybond N^+ membrane (Sambrook *et al*, 1989), and hybridized with cDNA probe at 42°C in Khandjian hybridizing buffer (Khandjian, 1989). The 3' part of wheat SBE I cDNA (designated wSBE I-D43, see Table 1) was labelled with the Rapid Multiprime DNA Probe Labelling Kit (Amersham) and used as probe. After washing at 60°C with 2 x SSC, 0.1% SDS three times, each time for about 1 to 2 hours, the membrane was visualized by overnight exposure at -80°C with X-ray film, Kodak MR.

<u>Example 2</u> <u>Frequency of Recovery of SBE I Type Clones from the Genomic Library</u>

An estimated 2 x 10^6 plaques from the amplified library were screened using an *Eco*RI fragment that contained 1200 bp at the 5′ end of maize SBE I (Baba *et al*, 1991) and twelve independent isolates were recovered and purified. This corresponds to the screening of somewhat fewer than the 2 x 10^6 primary plaques that exist in the original library (each of which has an average insert size of 15 kb) (Maniatis *et al*, 1982), because the amplification may lead to the representation of some sequences more than others. Assuming that the amplified library contains approximately three genomes of *T. tauschii*, the frequency with which SBE I-positive clones were recovered suggests the existence of about 5 copies of SBE I type genes within the *T. tauschii* genome.

Digestion of DNA from the twelve independent isolates by the restriction endonuclease *Bam*HI followed by hybridisation with a maize SBE I clone, suggested that the

genomic clones could be separated into two broad classes (Figure 1). One class had 10 members and a representative from this class is the clone λE1 (Figure 1, lane 1); λE6 (Figure 1, lane 3) is a member of this class, but is missing the 5' end of the E1-SBE I gene because the SBE I gene is at the extremity of the cloned DNA. Further hybridisation studies at high stringency with the extreme 5' and 3' regions of the SBE I gene contained in λE1 suggested that the other clones contained either identical or very closely related genes.

The second family had two members, and of these clone λE7 (Figure 1, lane 4) was arbitrarily selected for further study. These two members did not hybridise to probes from the extreme 5' and 3' regions of the SBE I gene that were contained in λE1, indicating that they were a distinct sub-class.

The DNA from *T. tauschii* and the lambda clones λE1 and λE7 was digested with *Bam*HI and hybridised with fragment E1.1, as shown in Figure 2. This fragment contains sequences that are highly conserved (85% sequence identity over 0.3 kB between λE1 and λE7), corresponding to exons 3, 4 and 5 of the rice gene. The bands in the genomic DNA at 0.8 kb and 1.0 kb correspond to identical sized fragments from λE1 and λE7, as shown in Figure 2; these are fragments E1.1 and E7.8 of λE1 and λE7 genomic clones respectively. Thus the arrangement of genes in the genomic clones is unlikely to be an artefact of the cloning procedure. There are also bands in the genomic DNA of approximately 2.5 kb, 4.8 kb and 8 kb in size which are not found from the digestion of λE1 or λE7; these could represent genes such as the 5' sequences of wSBE I-D1 or wSBE I-D3; see below.

Example 3 Tandem Arrangement of SBE I Type Genes in the *T. tauschii* Genome

Basic restriction endonuclease maps for λE1 and λE7 are shown in Figure 3. The map was constructed by

performing a series of hybridisations of *Eco*RI or *Bam*HI digested DNA from λE1 or λE7. The probes used were the fragments generated from *Bam*HI digestion of the relevant clone. Confirmation of the maps was obtained by PCR analysis, using primers both within the insert and also from the arms of lambda itself. PCR was performed in 10 μl volume using reagents supplied by Perkin-Elmer. The primers were used at a concentration of 20 μM. The program used was 94°C, 2 min, 1 cycle, then 94°C, 30 sec; 55°C, 30 sec; 72°C, 1min for 36 cycles and then 72°C, 5 min; 25°C, 1 min.

Sequencing was performed on an ABI sequencer using the manufacturer's recommended protocols for both dye primer and dye terminator technologies. Deletions were carried out using the Erase-a-base kit from Promega.

Sequence analysis was carried out using the GCG version 7 package of computer programs (Devereaux *et al*, 1984).

The PCR products were also used as hybridisation probes. The positioning of the genes was derived from sequencing the ends of the *Bam*HI subclones and also from sequencing PCR products generated from primers based on the insert and the lambda arms. The results indicate that there is only a single copy of a SBE I type gene within λE1. However, it is clear that λE7 resulted from the cloning of a DNA fragment from within a tandem array of the SBE I type genes. Of the three genes in the clone, which are named as wSBE I-D1, wSBE I-D2 and wSBE I-D3); only the central one (wSBE I-D2) is complete.

<u>Example 4</u> <u>Construction and Screening of cDNA Library</u>

A wheat cDNA library was constructed from the cultivar Rosella using pooled RNA from endosperm at 8, 12, 18 and 20 days after anthesis.

The cDNA library was prepared from poly A^+ RNA that was extracted from developing wheat grains (cv. Rosella, a hexaploid soft wheat cultivar) at 8, 12, 15, 18, 21 and 30 days after anthesis. The RNA was pooled and used

to synthesise cDNA that was propagated in lambda ZapII (Stratagene).

The library was screened with a genomic fragment from λE7 encompassing exons 3, 4 and 5 (fragment E7.8 in Figure 3). A number of clones were isolated. Of these an apparently full-length clone appeared to encode an unusual type of cDNA for SBE I. This cDNA has been termed SBE I-D2 type cDNA. The putative protein product is compared with the maize SBE I and rice SBE I type deduced amino acid sequences in Figure 4. The main difference is that this putative protein product is shorter at the C-terminal end, with an estimated molecular size of approximately 74 kD compared with 85 kDa for rice SBE I (Kawasaki *et al*, 1993). Note that amino acids corresponding to exon 9 of rice are missing in SBE I-D2 type cDNA, but those corresponding to exon 10 are present. There are no amino acid residues corresponding to exons 11-14 of rice; furthermore, the sequence corresponding to the last 57 amino acids of SBE I-D2 type has no significant homology to the sequence of the rice gene.

We expressed SBE I-D2 type cDNA in *E. coli* in order to examine its function. The cDNA was expressed as a fusion protein with 22 N-terminal residues of β-galactosidase and two threonine residues followed by the SBE I-D2 cDNA sequence either in or out of frame. Although an expected product of about 75 kDa in size was produced from only the in-frame fusion, we could not detect any enzyme activity from crude extracts of *E. coli* protein. Furthermore the in-frame construct could not complement an *E. coli* strain with a defined deletion in glycogen branching, although other putative branching enzyme cDNAs have been shown to be functional by this assay (data not shown). It is therefore unclear whether the wSBE I-D2 gene in λE7 codes for an active enzyme *in vivo*.

Example 5 Gene Structure in E7

i. Sequence of wSBE I-D2

We sequenced 9.2 kb of DNA that contained wSBE I-D2. This corresponds to fragments 7.31, 7.8 and 7.18. Fragment 7.31 was sequenced in its entirety (4.1 kb), but the sequence of about 30 bases about 2 kb upstream of the start of the gene could not be obtained because it was composed entirely of Gs. Elevation of the temperature of sequencing did not overcome this problem. Fragments 7.8 (1 kb) and 7.18 (4 kb) were completely sequenced, and corresponded to 2 kb downstream of the last exon detected for this gene. It was clear that we had isolated a gene which was closely related (approximately 95% sequence identity) to the SBE I-D2 type cDNA referred to above, except that the last 200 bp at the 3′ end of the cDNA are not present. The wSBE I-D2 gene includes sequences corresponding to rice exon 11 which are not in the cDNA clone. In addition it does not have exons 9, 12, 13 or 14; these are also absent from the SBE I-D2 type cDNA. The first two exons show lower identity to the corresponding exons from rice (approximately 60%) (Kawasaki *et al*, 1993) than to the other exons (about 80%). A diagrammatic exon-intron structure of the wSBE I-D2 gene is indicated in Figure 5. The restriction map was confirmed by sequencing the PCR products that spanned fragments 7.18 and 7.8 and 7.8 and E7.31 (see Figure 3) respectively.

ii. Sequence of wSBE I-D3

This gene was not sequenced in detail, as the genomic clone did not extend far enough to include the 5′ end of the sequence. The sequence is of a SBE-I type. The orientation of the gene is evident from sequencing of the relevant *Bam*HI fragments, and was confirmed by sequence analysis of a PCR product generated using primers from the right arm of lambda and a primer from the middle of the gene. The sequence homology with wSBEI-D2 is about 80% over the regions examined. The 2 kb sequenced corresponded to

exons 5 and 6 of the rice gene; these sequences were obtained by sequencing the ends of fragments 7.5, 7.4 and 7.14 respectively, although the sequences from the left end of fragment 7.14 did not show any homology to the rice sequences. The gene does not appear to share the 3' end of SBE I-D2 type cDNA, as a probe from 500 bp at the 3' end of the cDNA (including sequences corresponding to exons 8 and 10 from rice) did not hybridise to fragment 7.14, although it hybridised to fragment 7.18.

iii. Sequence of wSBE I-D1

This gene was also not sequenced in detail, as it was clear that the genomic clone did not extend far enough to include the 5' sequences. Limited sequencing suggests that it is also a SBE I type gene. The orientation relative to the left arm of lambda was confirmed by sequencing a PCR product that used a primer from the left arm of lambda and one from the middle of the gene (as above). Its sequence homology with wSBE I-D2 ,D3 and D4 (see below) is about 75% in the region sequenced corresponding to a part of exon 4 of the rice gene.

Starch branching enzymes are members of the α-amylase protein family, and in a recent survey Svensson (1994) identified eight residues in this family that are invariant, seven in the catalytic site and a glycine in a short turn. Of the seven catalytic residues, four are changed in SBE I-D2 type. However, additional variation in the 'conserved' residues may come to light when more plant cDNAs for branching enzyme I are available for analysis. In addition, although exons 9, 11, 12, 13 and 14 from rice are not present in the SBE I-D2 type cDNA, comparison of the maize and rice SBE I sequences indicate that the 3' region (from amino acid residue 730 of maize) is much more variable than the 5' and central regions. The active sites of rice and maize SBE I sequences, as indicated by Svensson (1994), are encoded by sequences that are in the central portion of the gene. When SBE II sequences from *Arabidopsis* were

compared by Fisher *et al* (1996) they also found variation at the 3′ and 5′ ends. SBE I-D2 type cDNA may encode a novel type of branching enzyme whose activity is not adequately detected in the current assays for detecting branching enzyme activity; alternatively the cDNA may correspond to an endosperm mRNA that does not produce a functional protein.

Example 6 Cloning of the cDNA corresponding to the wSBE I-D4 gene

The first strand cDNAs were synthesized from 1 µg of total RNA, derived from endosperm 12 days after pollination, as described by Sambrook *et al* (1989), and then used as templates to amplify two specific cDNA regions of wheat SBE I by PCR.

Two pairs of primers were used to obtain the cDNA clones BED1 and BED3 (Table 1). Primers used for cloning of BED3 were the degenerate primer NTS5'

5' GGC NAC NGC NGA G/AGA C/TGG 3' (SEQ ID NO.1),

based on the N-terminal sequence of the purified wheat endosperm SBE I protein, in which the 5' end of the primer is at position 168 of wSBE I-D4 cDNA, as shown in Table 1, based on the N-terminal sequence of wheat SBE I, and the primer NTS3'.

5' TAC ATT TCC TTG TCC ATCA 3' (SEQ ID NO.2)

in which the 5' end is at position 1590 of wSBE I-D4 cDNA, (see Table 1), designed to anneal to the conserved regions of the nucleotide sequences of BED5 and the maize and rice SBE I cDNAs. For clone BED1, the primers used were BEC5'

5' ATC ACG AGA GCT TGC TCA (SEQ ID NO.3)

in which the 5' end is at position 1 of wSBE I-D4 cDNA (see Table 1); the sequence was based on the wSBE I-D4 gene, and BEC3'

5' CGG TAC ACA GTT GCG TCA TTT TC 3' (SEQ ID NO.4)

in which the 5' end is at position 334 of wSBE I-D4 cDNA (see Table 1), and the sequence was based on BED 3.

Example 7 Identification of the gene from the *Triticum tauschii* SBE I family which is expressed in the endosperm

We have isolated two classes of SBE I genomic clones from *T. tauschii*. One class contained two genomic clone isolates, and this class has been characterised in some detail (Rahman *et al*, 1997). The complete gene contained within this class of clones was termed *wSBE I-D2*; there were additional genes at either ends of the clone, and these were designated *wSBE I-D1* and *wSBE I-D3*. The other class contained nine genomic clone isolates. Of these λE1 was arbitrarily taken as a representative clone, and its restriction map is shown in Figure 3; the SBE I gene contained in this clone was called *wSBE I-D4*.

Fragments E1.1 (0.8 kb) and E1.2 (2.1 kb) and fragments E1.7 (4.8 kb) and E1.5 (3 kb) respectively were completely sequenced. Fragment E1.7 was found to encode the N-terminal of the SBE I, which is found in the endosperm as described in Morell *et al* (1997). This is shown in Figure 6. Using antibodies raised against the N-terminal sequence, Morell *et al* (1997) found that the D genome isoform was the most highly expressed in the cultivars Rosella and Chinese Spring. We have thus isolated from *T. tauschii* a gene, *wSBE I-D4*, whose homologue in the hexaploid wheat genome encodes the major isoform for SBE I that is found in the wheat endosperm.

Table 1
Location of structural features and probes within wSBE I-D4 sequence.

A. Location of exons by comparison with the cDNA sequence of Repellin et al., (1997). Accession number Y12320.

Exon number	Start posn	End posn
1	4890	4987
2	5082	5149
3	5524	5731
4	5819	5888
5	6149	6318
6	6519	7424
7	7744	7860
8	8015	8077
9	8562	8670
10	9137	9237
11	9421	9488
12	9580	9661
13	9781	9897
14	9990	10480

B. Other features.

Name of feature.	wSBE I-D4. sequence	D4 cDNA sequence.
Putative initiation of translation	4900	11
Mature N-terminal sequence of SBE I	5550	124
End of translated SBE I sequence	10225	2431
End of D4 cDNA sequence	10461	2687
wSBE I-D45	4870,5860	1,354
wSBE I-D43	10116,10435	2338,2657
E1.1	5680,6400	380,630
BED 1		1,354
BED 2		169,418
BED 3		151,1601
BED 4		867,2372
BED 5		867,2687
Endosperm box like motif TGAAAAGT	4480,590	
CAAAT motif	4863	
TATAAA motif	4833	

All nine genomic clones of the λE1 type isolated from *T. tauschii* appear to contain the *wSBE I-D4* gene, or very similar genes, on the basis of PCR amplification and hybridisation experiments. However, the restriction patterns obtained for the clones differ with *Bam*HI and *Eco*RI, among other enzymes, indicating that either the clones represent near-identical but distinct genes or they represent the same gene isolated in distinct products of the *Sau*3A digest used to generate the library.

Example 8 Investigation of other SBE I genomic clones isolated

All ten members of the λE1-like class of SBE I genomic clones were investigated by hybridisation with probes derived from fragment E1.7 (sequence *wSBE I-D45*, encoding the translation start signal and the first 100 amino acids from the N-terminal end and intron sequences; see Table 1) and from fragment E1.5 (sequence *wSBE I-D43*, corresponding largely to the 3' untranslated sequence and containing intron sequences, see Table 1). The results obtained were consistent with one type of gene being isolated in different fragments in the different clones, as shown in Figure 7. The PCR products were obtained from the clones λE1, 2, 9, 14, 27, 31 and 52. These hybridised to *wSBE I-D45* using primers that amplify near the 5' end of the gene (positions 5590-6162 of *wSBE I-D4*). Sequencing showed no differences in sequence of a 200 bp product.

Analysis of the promoter for *wSBE I-D4* allows us to investigate the presence of motifs previously described for promoters that regulate gene expression in the endosperm. Forde *et al* (1985) compared prolamin promoters, and suggested that the presence of a motif approximately -300 bp upstream of the transcription start point, called the endosperm box, was responsible for endosperm-specific expression. The endosperm box was subsequently considered to consist of two different motifs: the endosperm motif (EM) (canonical sequence TGTAAAG) and the GCN 4 motif (canonical

sequence G/ATGAG/CTCAT). The GCN4 box is considered to regulate expression according to nitrogen availability (Muller and Knudsen, 1993). The *wSBE I-D4* promoter contains a number of imperfect EM-like motifs at approximately -100, -300 and -400 as well as further upstream. However, no GCN4 motifs could be found, which lends support to the idea that this motif regulates response to nitrogen, as starch biosynthesis is not as directly dependent on the nitrogen status of the plant as storage protein synthesis. Comparison of the promoters for *wSBE I-D4* and *D2* (Rahman *et al*, 1997) indicates that although there are no extensive sequence homologies there is a region of about 100 bp immediately before the first encoded methionine where the homology is 61% between the two promoters. In particular there is an almost perfect match in the sequence over twenty base pairs CTCGTTGCTTCC/TACTCCACT, (positions 4723-4742 of the *wSBE I* sequence), but the significance of this is hard to gauge, as it does not occur in the rice promoter for SBE I. The availability of more promoters for starch biosynthetic enzymes may allow firmer conclusions to be drawn. There are putative CAAT and TATA motifs at positions 4870 and 4830 respectively of *wSBE I-D4* sequence. The putative start of translation of the mRNA is at position 4900 of *wSBE I-D4*.

Figure 5 shows the structure of the *wSBE I-D4* gene, compared with the genes from rice and wheat (Kawasaki *et al*, 1993; Rahman *et al*, 1997). The rice *SBE I* has 14 exons compared with 13 for *wSBE I-D4* and 10 for *wSBE I-D2*. There is good conservation of exon-intron structure between the three genes, except at the extreme 5' end. In particular the sizes of intron 1 and intron 2 are very different between rice *SBE I* and *wSBE I-D4*.

<u>Example 9</u> <u>Isolation of cDNA for SBE I</u>

Using the maize starch branching enzyme I cDNA as a probe (Baba *et al*, 1991), 10 positive plaques were recovered by screening approximately 10^5 plaques from a wheat endosperm cDNA library prepared from the cultivar

Rosella, as described in Example 4. On purifying and sequencing these plaques it was clear that even the longest clone (BED5, 1822 bp) did not encode the N-terminal sequence obtained from protein analysis. Degenerate primers based on the wheat endosperm SBE I protein N-terminal sequence (Morell *et al*, 1997) and the sequence from BED5 were then used to amplify the 5' region: this produced a cDNA clone termed BED 3 (Table 1 and Figure 8). This cDNA clone overlapped extensively and had 100% sequence identity with BED5 and BED4 (Figure 8). As almost the entire protein N-terminal sequence had been included in the primer sequence design, this did not provide independent evidence of the selection of a cDNA sequence in the endosperm that encoded the protein sequence of the main form of SBE I. Using a BED3 to screen a second cDNA library produced BED2, which is shorter than BED3 but confirmed the BED3 sequence at 100% identity between positions 169 and 418 (Figure 8 and Table 1). In addition the entire cDNA sequence for BED3 could be detected at a 100% match in the genomic clone λE1. Primers based on the putative transcription start point combined with a primer based on the incomplete cDNAs recovered were then used to obtain a PCR product from total endosperm RNA by reverse transcription. This led to the isolation of the cDNA clone, BED1, of 300 bp, whose location is shown in Figure 8. By analysing this product, a sequence was again obtained that could be found exactly in the genomic clone λE1, and which overlapped precisely with BED3.

The N-terminal of the protein matches that of SBE I isolated from wheat endosperm by Morell *et al* (1997), and thus the *wSBE I-D4* cDNA represents the gene for the predominant SBE I isoform expressed in the endosperm. The encoded protein is 87 kDa; this is similar to proteins encoded by maize (Baba *et al*, 1991) and rice (Nakamura *et al*, 1992) cDNAs for SBE I and is distinct from the *wSBE I-D2* cDNA described previously, in which the encoded protein was 74 kDa (Rahman *et al*, 1997).

Five cDNA clones were sequenced and their sequences were assembled into one contiguous sequence using a GCG program (Devereaux *et al*, 1984). The arrangement of these sequences is illustrated in Figure 8, the nucleotide sequence is shown in SEQ ID No:5, and the deduced amino acid sequence is shown in SEQ ID No:6. The intact cDNA sequence, *wSBE I-D4* cDNA, is 2687 bp and contains one large open reading frame (ORF), which starts at nucleotides 11 to 13 and ends at nucleotides 2432 to 2434. It encodes a polypeptide of 807 amino acids with a molecular weight of 87 kDa. Comparison of the amino acid sequence encoded by *wSBE I-D4* cDNA with that encoded by maize and rice *SBE I* cDNAs showed that there is 75-80% identity between any of two these sequences at the nucleotide level and almost 90% at the amino acid level. Alignment of these three polypeptide sequences, as shown in Figure 4, along with the deduced sequences for pea, potato and *wSBE I-D2* type cDNA, indicated that the sequences in the central region are highly conserved, and sequences at the 5' end (about 80 amino acids) and the 3' end (about 60 amino acids) are variable.

Svensson *et al* (1994) indicated that there were several invariant residues in sequences of the α-amylase super-family of proteins to which SBE I belongs. In the sequence of maize SBE I these are in motifs commencing at amino acid residue positions 341, 415, 472, 537 respectively; these are also encoded in the *wSBE I-D4* sequence (SEQ ID No:9), further supporting the view that this gene encodes a functional enzyme. This is in contrast to the results with the *wSBE I-D2* gene, where three of the conserved motifs appear not to be encoded (Rahman *et al*, 1997).

There is about 90% sequence identity in the deduced amino acid sequence between *wSBE I-D4* cDNA and rice *SBE I* cDNA in the central portion of the molecule (between residues 160 and 740 for the deduced amino acid product from *wSBE I-D4* cDNA). The sequence identity of the deduced amino

acid sequence of the *wSBE I-D4* cDNA to the deduced amino acid sequence of *wSBE I-D2* is somewhat lower (85% for the most conserved region, between residues 285 to 390 for the deduced product of *wSBE I-D4* cDNA). Surprisingly, however, *wSBE I-D4* cDNA is missing the sequence that encodes amino acids at positions 30 to 58 in rice SBE I (see Figure 4). This corresponds to residues within the transit peptide of rice SBE I. A corresponding sequence also occurs in the deduced amino acid sequence from maize *SBE I* (Baba *et al*, 1991) and *wSBE I-D2* type cDNA (Rahman *et al*, 1997). Consequently the transit sequence encoded by *wSBE I-D4* cDNA is unusally short, containing only 38 amino acids, compared with 55-60 amino acids deduced for most starch biosynthetic enzymes in cereals (see for example Ainsworth, 1993; Nair *et al*, 1997). The *wSBE I-D4* gene does contain this sequence, but this does not appear to be transcribed into the major species of RNA from this gene, although it can be detected at low relative abundance. This raises the possibility of alternative splicing of the *wSBE I-D4* transcript, and also the question of the relative efficiency of translation/transport of the two isoforms. The possibility of alternative splicing in both rice and wheat has been considered for soluble starch synthase (Baba *et al*,1993 Rahman *et al*, 1995). Alternative splicing of soluble starch synthase would give a transit sequence of 40 amino acids, which is the same length proposed for the product of *wSBE I-D4* cDNA.

We have previously used probes based on exons 4, 5 and 6 (E7.8 and E1.1, see Rahman *et al.*, 1997) of *wSBE-D2* to probe wheat and *T. tauschii* genomic DNA cleaved with *Pvu*II and *Bam*HI respectively. This region is highly conserved within rice *SBE I*, *wSBE I-D2* and *wSBE I-D4* and produced ten bands with wheat DNA and five with *T. tauschii* DNA. Neither *Pvu*II nor *Bam*HI cleaved within the probe sequences, suggesting that each band represented a single type of SBE I gene. We have described four SBE I genes from *T. tauschii*: *wSBE I-D1*, *wSBE I-D2*, *wSBE I-D3* and *wSBE I-D4* (Rahman *et al*,

1997 and this specification), and so we may have accounted for most of the genes in *T. tauschii* and, by extension, the genes from the D genome of wheat. In wheat, at least two hybridising bands could be assigned to each of chromosomes 7A, 7B and 7D.

Example 10 Tissue specificity and expression during endosperm development

The 300 bp of 3' untranslated sequence of *wSBE I-D4* cDNA does not show any homology with either the *wSBE I-D2* type cDNA that we have described earlier (Rahman *et al*, 1997) or with BE-I from rice, as shown in Figure 5. We have called this sequence *wSBE I-D43C* (see SEQ ID No:9). It seemed likely that *wSBE I-D43C* would be a specific probe for this class of SBE-I, and thus it was used to investigate the tissue specificity. Hybridization of RNA from endosperm of hexaploid *T. tauschii* cultures with SBE I, SBE II, SSS I, DBE I, wheat actin, and wheat ribosomal RNA was examined. RNA was purified at various numbers of days after anthesis from plants grown with a 16 h photoperiod at 13 °C (night) and 18 °C (day). The age of the endosperms from which RNA was extracted in days after anthesis is given above the lanes in the blot. Equivalent amounts of RNA were electrophoresed in each lane. The probes used are identified in Tables 1 and 2.

The results are shown in Figures 9a to 9g. An RNA species of about 2700 bases in size was found to hybridise. This is very close to the size of the *wSBE I-D4* cDNA sequence. RNA hybridising to wSBE-I-D43C is most abundant at the mid-stage of endosperm development, as shown in Figure 9a, and in field grown material is relatively constant during the period 12-18 days, the time at which there is rapid starch and storage protein accummulation (Morell *et al*, 1995).

The sequence contained within the *wSBE I-D4* gene appears to be expressed only in the endosperm (Figure 9a, Figure 9b). We could not detect any expression in the leaf.

This could be because another isoform is expressed in the leaf, and/or because the amount of SBE I present in the leaf is much less than what is required in the endosperm. Isolation of SBE I clones from a leaf cDNA library would enable this question to be resolved.

Example 11 Intron-Exon Structure of SBE I

By comparison of the cDNA sequence of SBE I (Repellin *et al*, 1997) with that of *wSBE I-D4* we can deduce the intron-exon structure of the gene for the major isoform of SBE I that is found in the endosperm. The structure contains 14 exons compared to 14 for rice (Kawasaki *et al*, 1993). These 14 exons are spread over 6 kb of sequence, a distance similar to that found in both rice *SBE I* and *wSBE I-D2*. A dotplot comparison of *wSBE I-D4* sequence and that of rice *SBE I* sequence, depicted in Figure 10, shows good sequence identity over almost the entire gene starting from about position 5100 of *wSBE I-D4*; the identity is poor over the first 5 kb of sequence corresponding largely to the promoter sequences. The sequence identity over introns (about 60%) is lower than over exons (about 85%).

Example 12 Repeated Sequences in SBE I

Sequencing of *wSBE I-D4* revealed there was a repeated sequence of at least 300 bp contained in a 2kb fragment about 600 bp after the 3' end of the gene. We have called this sequence *wSBE I-D4R* (SEQ ID NO: 9). This repeated sequence is within fragment E1.5 (Figure 3 and Table 1) and is flanked by non-repetitive sequences from the genomic clone. We have previously shown that the restriction pattern obtained by digesting λE1 with the restriction enzyme *Bam*HI is also obtained when *T. tauschii* DNA is digested. Thus *wSBE I-D4R* is unlikely to be a cloning artefact. A search of the GenBank Database revealed that *wSBE I-D4R* shared no significant homology with any sequence in the database. Hybridisation experiments with *wSBE I-D4R* showed that all of the other *SBE I-D4* type

genomic clones (except number 29) contained this repeated sequence (data not shown). The *wSBE I-D4R* sequence was not highly repeated and occurred in the wheat genome with a similar frequency as the w*SBE I-D4* sequence.

When *SBE I-D4R* was used as the probe on wheat DNA from the nulli-tetra lines, four bands were obtained; two of these bands could be assigned to chromosome 7A and the others to chromosomes 7B and 7D (Figure 11). One of the two *Bam*HI fragments from wheat DNA which could be assigned to chromosome 7A was distinct from the single band from chromosome 7A detected using *wSBE I-D43* as the probe; the other three bands coincided in the autoradiograph with bands obtained with *wSBE I-D43*, and are likely to represent the same fragment. However, one of these fragments was distinct from the *Bam*HI fragment that hybridised to the *wSBE I-D43* sequence. In *wSBE I-D4* (see SEQ ID No:9), the *wSBE I-D43* sequence is only 300 bp upstream of *wSBE I-D4R*, and occurs in the same *Bam*HI fragment. These results suggest that the *wSBE I-D4R* sequence can occur independently of *wSBE I-D4* in the wheat genome.

Example 13 Isolation of Genomic Clones Encoding SBE II

Screening of a cDNA library, prepared from the wheat endosperm as described in Example 4, with the maize BE I clone (Baba *et al*, 1991) at low stringency led to the isolation of two classes of positive plaques. One class was strongly hybridising, and led to the isolation of wheat SBE I-D2 type and SBE I-D4 type cDNA clones, as described in Example 5 and in Rahman *et al* (1997). The second class was weakly hybridising, and one member of this class was purified. This weakly hybridising clone was termed SBE-9, and on sequencing was found to contain a sequence that was distinct from that for SBE I. This sequence showed greatest homology to maize BE II sequences, and was considered to encode part of the wheat SBE II sequence.

The screening of approximately 5 x 10^5 plaques from a genomic library constructed from *T. tauschii* (see

Example 1) with the SBE-9 sequence led to the isolation of four plaques that were positive. These were designated *wSBE II*-D1 to *wSBE II-D4* respectively, and were purified and analysed by restriction mapping. Although they all had different hybridization patterns with SBE-9, as shown in Figure 12, the results were consistent with the isolation of the same gene in different-sized fragments.

Example 14 Identification of the N-terminal sequence of SBE II

Sequencing of the SBE II gene contained in clone 2, termed *SBE II-D1* (see SEQ ID No:10), showed that it coded for the N-terminal sequence of the major isoform of SBE II expressed in the wheat endosperm, as identified by Morell *et al* (1997). This is shown in Figure 13.

Example 15 Intron-Exon Structure of the SBE II Gene

In addition to encoding the N-terminal sequence of sBE II, as shown in Example 10, the cDNA sequence reported by Nair *et al* (1997) was also found to have 100% sequence identity with part of the sequence of *wSBE II-D1*. Thus the intron-exon structure can be deduced, and this is shown in Figure 14. The positions of exons and other major structural features of the SBE II gene are summarized in Table 2.

Example 16 Number of SBE II Genes in *T. tauschii* and Wheat

Hybridisation of the SBE II conserved region with *T. tauschii* DNA revealed the presence of three gene classes. However, in our screening we only recovered one class. Hybridisation to wheat DNA indicated that the locus for SBE II was on chromosome 2, with approximately 5 loci in wheat; most of these appear to be on chromosome 2D, as shown in Figure 15.

Table 2
Positions of structural features in wSBE II-D1.

A. Positions of exons.

Exon number	Genomic start	Genomic finish
1	1058	1336
2	1664	1761
3	2038	2279
4	2681	2779
5	2949	2997
6	3145	3204
7	3540	3620
8	3704	3825
9	4110	4188
10	4818	4939
11	5115	5234
12	6209	6338
13	6427	6549
14	6739	6867
15	7447	7550
16	8392	8536
17	9556	9703
18	9839	9943
19	10120	10193
20	10395	10550
21	10928	11002
22	11092	11475

B. Other structural features within the wSBE II-D1 DNA sequence

Putative initiation of translation	1214
Mature N-terminal sequence of SBE II.	1681
wSBE II-D13	11116 to 11448
Endosperm box like motif TGAAAAGT	521
Endosperm box like motif TGAAAGT	565
Endpsperm box like motif CGAAAAT	669
Endosperm box like motif TAAATGT	768
CAAAAT motif	784
TCAATT motif	1108
TATAAA motif	799
AATTAA motif	1110

Example 17 Expression of SBE II

Investigation of the pattern of expression of SBE II revealed that the gene was only expressed in the endosperm. However the timing of expression was quite distinct from that of SBE I, as illustrated in Figures 9a, 9b and 9c.

SBE I gene expression is only clearly detectable from the mid-stage of endosperm development (10 days after anthesis in Figure 9b), whereas SBE II gene expression is clearly seen much earlier, in endosperm tissue at 5-8 days after development (Figures 9a and 9c), corresponding to an early stage of endosperm development. The hybridisation of wheat endosperm mRNA with the actin and ribosomal RNA genes is shown as controls (Figures 9fa and 9g, respectively).

Example 18 Cloning of Wheat Soluble Starch Synthase cDNA

A conserved sequence region was used for the synthesis of primers for amplification of SSS I by comparison with the nucleotide sequences encoding soluble starch synthases of rice and pea. A 300 bp RT-PCR product was obtained by amplification of cDNA from wheat endosperm at 12 days post anthesis. The 300 bp RT-PCT product was then cloned, and its sequence analysed. The comparison of its sequence with rice SSS cDNA showed about 80% sequence homology. The 300 bp RT-PCR product was 100% homologous to the partial sequence of a wheat SSS I in the database produced by Block *et al* (1997).

The 300 bp cDNA fragment of wheat soluble starch synthase thus isolated was used as a probe for the screening of a wheat endosperm cDNA library (Rahman *et al*, 1997). Eight cDNA clones were selected. One of the largest cDNA clones (sm2) was used for DNA sequencing analysis, and gave a 2662 bp nucleotide sequence, which is shown in SEQ ID NO:14. A large open reading frame of this cDNA encoded a 647 amino acid polypeptide, starting at nucleotides 247 to 250 and terminating at nucleotides 2198 to 2200. The

deduced polypeptide was shown by protein sequence analysis to contain the N-terminal sequence of a 75 kDa granule-bound protein (Rahman *et al*, 1995). This is illustrated in Figure 16. The location of the 75 kDa protein was determined for both the soluble fraction and starch granule-bound fraction by the method of Denyer *et al* (1995). Thus this cDNA clone encoded a polypeptide comprising a 41 amino acid transit peptide and a 606 amino acid mature peptide (SEQ ID NO:12). The cleavage site LRRL was located at amino acids 36 to 39 of the transit peptide of this deduced polypeptide.

Comparison of wheat SSS I with rice SSS and potato SSS showed that there is 87.4% or 75.9% homology at the amino acid level and 74.7% or 58.1% homology at the nucleotide level. Some amino acids in the at N-terminal sequences of the SSS I of wheat and rice were conserved. Major features of the SSS I gene are summarized in Table 3.

<u>Example 19</u> <u>Isolation of Genomic Clone of Wheat Soluble Starch Synthase</u>

Seven genomic clones were obtained with a 300 bp cDNA probe by screening approximately 5 x 10^5 plaques from a genomic DNA library of *Triticum tauschii*, as described above. DNA was purified from 5 of these clones and digested with *Bam*HI and *Sac*I. Southern hybridization analysis using the 300 bp cDNA as probe showed that these clones could be classified into two classes, as shown in Figure 17. One genomic clone, sg3, contained a long insert, and was digested with *Bam*HI or *Sac*I and subcloned into pBluescript KS+ vector.

Table 3

Comparison of exons and introns of soluble starch synthases I genes of wheat and rice

(1) Identity of exons of soluble starch synthase I genes of wheat and rice

Exons	wSSI-D1	rSSI	identity (%)	start site (wSSI-D1)	stop site (wSSI-D1)
1a	255	113	57.52	-253	0
1b	316	298	58.92	1	316
2	356	356	82.87	1473	1828
3	78	78	92.31	2746	2823
4	125	125	90.40	2906	3028
5	82	82	89.02	4113	4194
6	174	174	93.10	4286	4459
7	82	82	93.90	4562	4643
8	92	92	92.39	4743	4835
9	63	63	90.48	4959	5021
10	90	90	82.22	5103	5192
11	125	125	88.80	8594	8718
12	109	109	91.74	8807	8915
13	53	53	81.13	8992	9044
14	40	41	80.00	9160	9199
15a	159	113	79.65	9499	9657
15b	392	539	46.46	9658	10098

(2) Identity of introns of soluble starch synthase I genes of wheat and rice

Introns	wSSI-D1	rSSI	identity (%)	start site (wSSI-D1)	stop site (wSSI-D1)
1	1156	907	41.05	317	1472
2	917	851	41.65	1829	2745
3	82	87	45.12	2824	2905
4	1084	835	48.50	3029	4112
5	91	96	57.78	4195	4285
6	102	189	52.48	4460	4561
7	99	96	52.08	4644	4742
8	123	110	45.46	4836	4958
9	81	78	58.97	5022	5102
10	3401	663	37.56	5193	8593
11	88	124	56.82	8719	8806
12	76	81	48.68	8916	8991
13	115	135	45.22	9045	9159
14	299	830	45.80	9200	9498

Note: Exon 1a: non-coding region of exon 1. Exon 1b: coding region of exon 1.

Exon 15a: coding region of exon 15. Exon 15b: non-coding region of exon 15.

wSSI-D1: wheat soluble starch synthase I gene.

rSSI: rice soluble starch synthase I gene.

These subclones were analysed by sequencing. The intron/exon structure of the sg3 rice gene is shown in Figure 18. The SSS I gene from *T. tauschii* is shown in SEQ ID No:13, while the deduced amino acid sequence is shown in SEQ ID NO:14.

Example 20 Northern Hybridization Analysis of the Expression of Genes Encoding Soluble Starch Synthase

Total RNAs were purified from leaves, pre-anthesis material, and various stages of developing endosperm at 5-8, 10-15 and 18-22 days post anthesis. Northern hybridization analysis showed that mRNAs encoding wheat SSS I were specifically expressed in developmental endosperm. Expression of this mRNAs in the leaves and pre-anthesis materials could not be detected by northern hybridization analysis under this experimental condition. Wheat SSS I mRNAs started to express at high levels at an early stage of endosperm, 5-8 days post anthesis, and the expression level in endosperm at 10-15 days post anthesis, was reduced. These results are summarized in Figure 9a and Figure 9d.

Example 21 Genomic Localisation of Wheat Soluble Starch Synthase

DNA from chromosome engineered lines was digested with the restriction enzyme BamHI and blotted onto supported nitrocellulose membranes. A probe prepared from the 3' end of the cDNA sequence, from positions 2345 to 2548, was used to hybridise to this DNA. The presence of a specific band was shown to be associated with the presence of chromosomes 7A (Figure 19). These data demonstrate location of the SSS I gene on chromosome 7.

Example 22 Isolation of SSS I Promoter

We have isolated the promoter that drives this pattern of expression for SSS I. The pattern of expression for SSS I is very similar to that for SBE II: the SSS I gene

transcript is detectable from an early stage of endosperm development until the endosperm matures. The sequence of this promoter is given in SEQ ID No:15.

<u>Example 23</u> <u>Isolation of the Gene Encoding Debranching Enzyme from Wheat</u>

The *sugary*-1 mutation in maize results in mature dried kernels that have a glassy and translucent appearance; immature mature kernels accumulate sucrose and other simple sugars, as well as the water-soluble polysaccharide phytoglycogen (Black *et al*, 1966). Most data indicates that in *sugary-1* mutants the concentration of amylose is increased relative to that of amylopection. Analysis of a particular *sugary-1* mutation (*su-1Ref*) by James *et al*, (1995) led to the isolation of a cDNA that shared significant sequence identity with bacterial enzymes that hydrolyse the α 1,6-glucosyl linkages of starch, such as an isoamylase from *Pseudomonas* (Amemura *et al*, 1988), <u>ie</u>. bacterial debranching enzymes.

We have now isolated a sequence amplified from wheat endosperm cDNA using the polymerase chain reaction (PCR). This sequence is highly homologous to the sequence for the *sugary* gene isolated by James *et al*, (1995). This sequence has been used to isolate homologous cDNA sequences from a wheat endosperm library and genomic sequences from *Triticum tauschii*.

Comparison of the deduced amino acid sequences of DBE from maize with spinach (Accession SOPULSPO, GenBank database), *Pseudomonas* (Amemura *et al*, 1988) and rice (Nakamura *et al*, 1997) enabled us to deduce sequences which could be useful in wheat. When these sequences were used as PCR amplification primers with wheat genomic DNA a product of 256 bp was produced. This was sequenced and was compared to the sequence of maize *sugary* isolated by James *et al*, (1995). The results are shown in Figure 20a and Figure 20b. This sequence has been termed wheat debranching enzyme sequence I (WDBE-I).

WDBE-1 was used to investigate a cDNA library constructed from wheat endosperm (Rahman *et al*, 1997) enables us to isolate two cDNA clones which hybridise strongly to the WDBE-I probe. The nucleotide sequence of the DNA insert in the longest of these clones is given in SEQ ID No:16.

Use of WDBE 1 to investigate a genomic library constructed from *T. tauschii*, as described above has led to the isolation of four genomic clones, designated I1, I2, I3 and I4, respectively, which hybridised strongly to the WDBE-I sequence. These clones were shown to contain copies of a single debranching enzyme gene. The sequence of one of these clones, I2, is given in SEQ ID No:17. The intron/exon structure of the gene is shown in Figure 20c. Exons 1 to 4 were identified by comparison with the maize *sugary*-1 cDNA, while Exons 5 to 18 were identified by comparison with the cDNA sequence given in SEQ ID No:16. The major features of the DBE I gene are summarized in Table 4.

Hybridization of WDBE-I to DNA from *T. tauschii* indicates one hybridizing fragment (Figure 21a). The chromosomal location of the gene was shown to be on chromosome 7 through hybridisation to nullisomic/tetrasomic lines of the hexaploid wheat cultivar Chinese Spring (Figure 21b).

We have clearly isolated a sequence from the wheat genome that has high identity to the debranching enzyme cDNA of maize characterised by James *et al* (1997). The isolation of homologous cDNA sequences and genomic sequences enables further characterisation of the debranching enzyme cDNA and promoter sequences from wheat and *T. tauschii*. These sequences and the WDBE I sequences shown herein are useful in the manipulation of wheat starch structure through genetic manipulation and in the screening for mutants at the equivalent *sugary* locus in wheat.

Figure 9e shows that the DBE I gene is expressed during endosperm development in wheat and that the timing of expression is similar to the SBEII and SSSI genes. Figure 9h

shows that the full length mRNA for the gene (3.0 kb) is found only in the wheat endosperm.

Example 24 Transient assays of Promoter-GFP Fusions

DNA constructs

DNA constructs for transient expression assays were prepared by fusing sequences from the BEII and SSI promoters to the gene encoding the Green Fluorescent Protein. Green Fluorescent Protein (GFP) constructs contained the GFP gene described by Sheen et al. (1995). The nos 3′ element (Bevan et al., 1983) was inserted 3′ of the GFP gene. The plasmid vector (pWGEM_NZfp) was constructed by inserting the NotI to HindIII fragment from the following sequence:

5′ GCGGCCGCTC CCTGGCCGAC TTGGCCGAAG CTTGCATGCC TGCAGGTCGA CTCTAGAGGA TCCCCGGGTA CCGAGCTCGA ATTCATCGAT GATATCAGAT CCGGGCCCTC TAGATGCGGC CGCATGCATA AGCTT 3′

into the *Not*I and *Hind*III sites of pGem-13Zf(-) vector (Promega). The sequences at the junction of the wSSSIpro1 and wSSSIpro2 and GFP were identical, and included the junction sequence:

5′....CGCGCGCCCA CACCCTGCAG GTCGACTCTA GAGGATCCAT GGTGAGCAAG 3′.

The sequence at the junction of wsbeIIpro1 and GFP was:

5′ GCGACTGGCT GACTCAATCA CTACGCGGGG ATCCATGGTG AGCAAGGGCG 3′.

The sequence at the junction of wsbeIIpro2 and GFP was:

5′ GGACTCCTCT CGCGCCGTCC TGAGCCGCGG ATCCATGGTG AGCAAGGGCG 3′.

The structures of the constructs are shown in Figures 22a to 22f.

Table 4
Structural features of wDBEI-D1

A.
Position of exons

Exon number	Start position	End position	Comments
1	1890	2241	(deduced by comparison with maize)
2	2342	2524	(deduced by comparison with maize)
3	2615	2707	(deduced by comparison with maize)
4	3016	3168	(deduced by comparison with maize)
5	3360	3436	
6	4313	4454	
7	4526	4633	
8	4734	4819	
9	5058	5129	
10	5202	5328	
11	5558	5644	
12	6575	6671	
13	7507	7661	
14	8450	8527	
15	8739	8823	
16	8902	8981	
17	9114	9231	
18	Still being sequenced		

Note that following nucleotides 3330, 6330 and 8419 there may be short regions of DNA not yet sequenced.

B.

CAAAAT motif	1833
TCAAT motif	1838
ATAAATAA motif	1804
Endosperm box like motif TAAAACG	1463

Preparation of target tissue

All explants used for transient assay were from the hexaploid wheat cultivar, Milliwang. Endosperm (10 - 12 days after anthesis), embryos (12 - 14 days after anthesis) and leaves (the second leaf from the top of plants containing 5 leaves) were used. Developing seed or leaves were collected, surface sterilized with 1.25% w/v sodium hypochlorite for 20 minutes and rinsed with sterile distilled water 8 times. Endosperms or embryos were carefully excised from seed in order to avoid contamination with surrounding tissues. Leaves were cut into 0.5 cm x 1 cm pieces. All tissues were aseptically transferred onto SD1SM medium, which is an MS based medium containing 1 mg/L 2,4-D, 150 mg/L L-asparagine, 0.5 mg/L thiamine, 10 g/L sucrose, 36 g/L sorbitol and 36 g/L mannitol. Each agar plate contained either 12 endosperms, 12 embros or 2 leaf segments.

Preparation of gold particles and bombardment

Five µg of each plasmid was used for the preparation of gold particles, as described by Witrzens *et al.* (1998). Gold particle-DNA suspension in ethanol (10 µl) was used for each bombardment using a Bio-Rad helium-driven particle delivery system, PDS-1000.

GFP assay

The expression of GFP was observed after 36 to 72 hours incubation using a fluorescence microscope. Two plates were bombarded for each construct. The numbers of expressing regions were recorded for each target tissue, and are summarized in Table 5. The intensity of the expression of GFP from each of the promoters was estimated by visual comparison of the light intensity emitted, and is summarized in Table 6.

The DNA construct containing GFP without a promoter region (pZLGFPNot) gave no evidence of transient expression in embryo (panel l) or leaf (panel r) and

extremely weak and sporadic expression in endosperm (panel f) , this construct gave only very weak expression in endosperm with respect to the number (Figure 5) and intensity (Figure 6) of transient expression regions. The constructs pwsssIpro1gfpNOT (panels b, h and n), psbeIIpro1gfpNOT(panels d, j and p), and psbeIIpro2gfpNOT (panels e, k and q) yielded low numbers (Table 5) of strongly (Table 6) expressing regions in leaves, and there was a very uneven distribution of expressing regions between target leaf pieces (Table 5). pwsssIpro2gfpNOT (panels c, i and o) gave no evidence of transient expression in leaves (Table 5). These results show that each of the promoter constructs is able to drive the transient expression of GFP in the grain tissues, endosperm and embryo. The ability of the short SSI promoter (pwsssIpro2gfpNOT containing 1042 bp 5′ of the ATG translation start site) to drive expression in leaves (panel n) contrasts with the inability of the long SSI promoter (pwsssIpro2gfpNOT containing 3914 base pair region 5′ of the ATG translation start site, panel o)) suggesting that regions for controlling tissue specificity are located between -3914 and -1042 of the SSI promoter region (SEQ ID No:15).

Example 25 Stable transformation of rice

Stable transformation of rice using *Agrobacterium* was carried out essentially as described by Wang et al. 1997. The plasmids containing the target DNA constructs containing the promoter-reporter gene fusions are shown in Figure 23. These plasmids were transformed into *Agrobacterium tumefaciens* AGL1 by electroporation.and cultured on selection plates of LB media containing rifampicillin (50 mg/L) and spectinomycin (50 mg/L) for 2 to 3 days, and then gently suspended in 10 ml NB liquid medium containing 100 µM acetosyringone and mixed well. Embryogenic rice calli (2 to 3 months old) derived from mature seeds were immersed in the *A. tumefaciens* AGL1

Table 5
Transient Assay of GFP based constructs

Tissue	Construct	Plate No.	Explant Number 1	2	3	4	5	6	7	8	9	10	11	12	Ave.	S.D.
Endosperm	pact_jsgfg_nos	1	0	0	1	158	152	148	0	2	12	159	95	64	65.9	71.6
Endosperm	pact_jsgfg_nos	2	3	13	2	83	18	9	6	188	0	102	5	3	36.0	58.6
Embryo	pact_jsgfg_nos	3	97	79	77	101	121	176	89	129	139	212	131	138	124.1	40.1
Embryo	pact_jsgfg_nos	4	18	39	89	82	7	52	94	147	19	66	106	85	67.0	41.6
Leaf	pact_jsgfg_nos	5	0	2	0	3	0	0							0.8	1.3
Leaf	pact_jsgfg_nos	6	0	0	0	1	0	0							0.2	0.4
Leaf	pact_jsgfg_nos	7	3	0	0	2	0	3							1.3	1.5
Endosperm	pZLGFPNot	8	13	0	4	0	14	0	0	0	0	0	0	1	2.7	5.2
Endosperm	pZLGFPNot	9	0	0	0	0	14	0	0	5	3	4	6	0	2.7	4.2
Embryo	pZLGFPNot	10	0	0	0	0	0	0	0	0	0	0	0	0	0.0	0.0
Embryo	pZLGFPNot	11	0	0	0	0	0	0	0	0	0	0	0	0	0.0	0.0
Leaf	pZLGFPNot	12	0	0	0	0	0	0							0.0	0.0
Leaf	pZLGFPNot	13	0	0	0	0	0	0							0.0	0.0
Leaf	pZLGFPNot	14	0	0	0	0	0	0							0.0	0.0

Table 5 (Continued)
Transient Assay of GFP based constructs

Tissue	Construct	Plate No.	Explant Number												Ave.	S.D.
Endosperm	psbeIIpro1gfpNOT	15	111	0	77	142	0	127	7	35	39	191	95	34	71.5	62.3
Endosperm	psbeIIpro1gfpNOT	16	21	101	0	0	34	164	102	5	39	125	147	114	71.0	60.6
Embryo	psbeIIpro1gfpNOT	17	23	67	63	4	12	14	9	8	29	19	24	51	26.9	21.7
Embryo	psbeIIpro1gfpNOT	18	92	144	64	36	31	23	106	43	11	1	9	7	47.3	45.4
Leaf	psbeIIpro1gfpNOT	19	0	0	0	0	0	0							0.0	0.0
Leaf	psbeIIpro1gfpNOT	20	6	0	0	0	0	0							1.0	2.4
Leaf	psbeIIpro1gfpNOT	21	0	0	0	0	3	5							1.3	2.2
Endosperm	psbeIIpro2fpNOT	22	12	18	3	0	0	21	13	0	10	11	10	0	8.2	7.4
Endosperm	psbeIIpro2fpNOT	23	24	25	13	68	11	0	0	0	1	0	0	0	11.8	20.1
Embryo	psbeIIpro2fpNOT	24	9	13	4	7	6	21	0	9	3	5	2	4	6.9	5.7
Embryo	psbeIIpro2fpNOT	25	5	0	3	5	23	4	3	1	8	12	8	13	7.1	6.4
Leaf	psbeIIpro2fpNOT	26	0	2	0	0	0	0							0.3	0.8
Leaf	psbeIIpro2fpNOT	27	0	5	0	8	0	0							2.2	3.5
Leaf	psbeIIpro2fpNOT	28	0	0	0	0	0	0							0.0	0.0

Table 5(Continued)
Transient Assay of GFP based constructs

Tissue	Construct	Plate No.	Explant Number												Ave.	S.D.
Endosperm	pwsssIpro1gfpNOT	29	121	0	0	28	0	4	81	23	0	2	0	2	21.8	39.2
Endosperm	pwsssIpro1gfpNOT	30	3	0	0	92	12	0	0	102	4	159	41	24	36.4	52.8
Embryo	pwsssIpro1gfpNOT	31	112	106	74	54	33	73	77	49	42	38	59	46	63.6	25.6
Embryo	pwsssIpro1gfpNOT	32	97	48	110	22	191	112	53	6	9	145	6	10	67.4	62.4
Leaf	pwsssIpro1gfpNOT	33	0	0	0	0	0	0							0.0	0.0
Leaf	pwsssIpro1gfpNOT	34	0	0	0	0	0	0							0.0	0.0
Leaf	pwsssIpro1gfpNOT	35	12	0	0	0	0	0							2.0	4.9
Endosperm	pwsssIpro2fpNOT	36	0	0	18	81	0	0	0	6	0	0	1	0	8.8	23.3
Endosperm	pwsssIpro2fpNOT	37	0	18	14	6	63	8	8	23	79	7	46	51	26.9	26.1
Embryo	pwsssIpro2fpNOT	38	15	7	14	57	8	3	26	10	47	34	47	0	22.3	19.4
Embryo	pwsssIpro2fpNOT	39	9	15	48	103	31	22	107	22	27	82	51	63	48.3	33.8
Leaf	pwsssIpro2fpNOT	40	0	0	0	0	0	0							0.0	0.0
Leaf	pwsssIpro2fpNOT	41	0	0	0	0	0	0							0.0	0.0
Leaf	pwsssIpro2fpNOT	42	0	0	0	0	0	0							0.0	0.0

Table 6
Comparison of the Intensities of Transient Expression

Tissue	pact_js-gfg_nos	pwsssI-pro1gfpNOT	pwsssI-pro2gfpNOT	psbeII-pro1gfpNOT	psbeII-pro2gfpNOT	pZLGFP Not
Endosperm	10	4	2.5	3.5	1.5	0.5
Embryo	10	5.5	5.5	1.5	1	0
Leaf	10	20	0	10	10	0

All intensities are relative to pact_js-gfg_nos transient expression in the target tissue
Relative intensities were independently scored by three researchers and averaged.

suspension. After 3 - 10 minutes the *A. tumefaciens* AGL1 suspension medium was removed, and the rice calli were transferred to NB medium containing 100 μM acetosyringone for 48 h. The co-cultivated calli were washed with sterile Milli Q H_2O containing 150 mg/L timentin 7 times to remove all *Agrobacterium*, plated on to NB medium containing 150 mg/L timentin and 30 mg/L hygromycin, and cultured for 3 to 4 weeks. Newly-formed buds on the surface of rice calli were excised and plated onto NB Second Selection medium containing 150 mg/L timentin and 50 mg/L hygromycin. After 4 weeks of proliferation calli were plated onto NB Pre-Regeneration medium containing 150 mg/L timentin and 50 mg/L hygromycin, and cultured for 2 weeks. The calli were then transferred on to NB-Regeneration medium containing 150 mg/L timentin and 50 mg/L hygromycin for 3 to 4 weeks. Once shooting occurs, shoots are transferred onto rooting medium (½ MS) containing 50 mg /L hygromycin. Once adequate root formation occurs, the seedlings are transferred to soil, grown in a misting chamber for 1-2 weeks, and grown to maturity in a containment glasshouse.

Example 26 Use of probes from SSS I, SBE I, SBE II and DBE sequences to identify null or altered alleles for use in breeding programmes

DNA primer sets were designed to enable amplification of the first 9 introns of the SBE II gene using PCR. The design of the primer sets is illustrated in Figure 24. Primers were based on the wSBE II-D1 sequence (deduced from Figure 13b and Nair *et al*, 1997; Accession No. Y11282) and were designed such that intron sequences in the wSBE II sequence were amplified by PCR. These primer sets individually amplify the first 9 introns of SBE II. One primer (sr913F) contained a fluorescent label at the 5' end. Following amplification, the products were digested with the restriction enzyme Dde1 and analysed using an ABI 377 DNA Sequencer with Genescan™ fragment analysis software. One primer set, for intron 5, was found to amplify products from

each of chromosomes 2A, 2B and 2D of wheat. This is shown in Figure 25, which illustrates results obtained with various wheat lines, and demonstrates that products from each of the wheat genomes from diverse wheats were amplified, and that therefore lines lacking the wSBEII gene on a specific chromosome could be readily identified. Lane (iii) illustrates the identification of the absence of the A genome wSBEII gene from the hexaploid wheat cultivar Chinese Spring ditelosomic line 2AS.

Figure 26 compares results of amplification with an Intron 10 primer set for various nullisomic/tetrasomic lines of the hexaploid wheat Chinese Spring. Fluorescent dUTP deoxynucleotides were included in the amplification reaction. Following amplification, the products were digested with the restriction enzyme *Dde*I and analysed using an ABI 377 DNA Sequencer with Genescan™ fragment analysis software. In lane (i) Chinese Spring ditelosomic line 2AS, a 300 base product is absent; in lane (ii) N2BT2A, a 204 base product is absent, and in lane (iii) N2DT2B a 191 base product is absent. These results demonstrate that the absence of specific wSBEII genes on each of the wheat chromosomes can be detected by this assay. Lines lacking wSBEII forms can be used as a parental line for breeding programmes for generation of new lines in which expression of SBE II is diminished or abolished, with consequent increase in amylose content of the wheat grain. Thus a high amylose wheat can be produced.

Table 7 shows examples primers pairs for SBE I, SSS I and DBE I which can identify genes from individual wheat genomes and could therefore be used to identify lines containing null or altered alleles. Such tests could be used to enable the development of wheat lines carrying null mutations in each of the genomes for a specific gene (for

Table 7

PCR Primers for Starch Biosynthesis Genes

Gen	Foward Primer	Foward Primer sequence	Reverse Primer	Reverse Primer sequence	Temp (°C)	Product (bp)
SBE I	ZLE1 5d	GGC GGC GGC AAT GTG CGG CTG AG	ZLBE1 63	CCA GAT CGT ATA TCG GAA GGT CG	57.3	A=625, B = 600, D = 550
SSS I	sssE01F	GAA CTC GCG CCC GAC CTC CT	ZLSg7	AGC CAC GAT TAT GCT GTC GAT GG	55.0	A, 450; B=450; D= 630
	sssE14F	TTC TCA CCG CTA ACC GTG GAC	ZLSm19	GTC TAC ATG ACG TAG GGT TGG TC	55.8	B = 400, D = 500 no A product
DBE I	DBEE17F	TGG TCT GAG AAT AGC CGA TTC	sr1536F	AAGGCCACATAGATCTCG	56.8	B, 190; D, 190, A, 160. Non-specific product 220 bp

Temp: = annealing temperature, bp = length of the product in base pairs

example SBEI, SSI or DBE I) or combinations of null alleles for different genes.

It will be apparent to the person skilled in the art that while the invention has been described in some detail for the purposes of clarity and understanding, various modifications and alterations to the embodiments and methods described herein may be made without departing from the scope of the inventive concept disclosed in this specification.

Reference cited herein are listed on the following pages, and are incorporated herein by this reference.

REFERENCES

Ainsworth, C., Clark, J. and Balsdon, J.
Plant Molecular Biology, 1993 22 67-82

Amemura, A., Chakrabort, R., Fujita, M., Noumi, T. and Futai, M.
Biol. Chem., 1988 263 9271-9275

Baba, T., Kimura, K., Mizuno, K., Etoh, H., Ishida,Y., Shida, O. and Arai, Y.
Biochem. Biophys. Res. Commun., 1991 181 87-94.

Baba,T.; Nishihara,M.; Mizuno,K.; Kawasaki,T.; Shimada,H.; Kobayashi,E.; Ohnishi,S.; Tanaka,K.; Arai,Y.
Plant Physiol, 1993, 103 565-573.

Ball,S.; Guan,H.P.; James,M.; Myers,A.; Keeling,P.; Mouille,G.; Buléon,A.; Colonna,P.; Preiss,J.
Cell, 1996, 86 349-352

Bevan, M., Barnes, W.M., and Chilton, M.
Nucleic Acids Research, 1983, 11 369-385

Black, R.C., Loerch, J.D., McARdle, F.J. and Creech, R.G.
Genetics, 1966 53 661-668

Block, M., Loerz, H., Lutticke, S.
Genbank database Accession number U48227

Burton, R.A., Bewley, J.D., Smith, A.M., Bhattacharya, M.K., Tatge, H., Ring,S., Bull, V., Hamilton, W.D.O. and Martin, C.
The Plant Journal, 1995 7 3-15.

Cangiano, G., La Volpe, A., Poulsen, P. and Kreiberg, J.D.
Plant Physiology, 1993 102 1053-1054.

Clarke, B.C., Mukai, Y. and Appels, R.
Chromosoma, 1996 105 269-275

Devereaux, J., Haeberli, P. and Smithies, O.
Nucleic Acids Res., 1984 12, 387-395.

Denyer, K., Hylton, C.M., Jenner, C.F. and Smith, A.M.
Planta, 1995 196 256-265

Doherty, J.P., Lindeman, R., Trent, R.J., Graham, M.W. and Woodcock, D.M.
Gene, 1992 124 113-120

Dry, I., Smith, A., Edwards, A., Bhattacharyya, M., Dunn, P., Martin, C.
Plant J 1992, 2 193-202
Edwards, A., Marshall, J., Sidebottom, C., Visser, R.G.F., Smith, A.M., Martin, C.
Plant J, 1995 8 283-294
Fisher, D.K., Boyer, C.D. and Hannah, L.C.
Plant Physiology, 1993 102 1045-1046
Forde, B.G., Heyworth, A., Pywell, J. and Forde, M.
Nucleic Acids Research, 1985 13 7327-7339
Gill, B.S. and Appels, R.
Plant Syst. Evol., 1988 160 77-90.
Higgins, T.J.V., Zwar, J.A., Jacobsen, J.V. (1976)
Nature, 1976, 260 166-168
Khandjian, E.W.
Bio/Technology, 1987, 5 165-167
Jahne, A., Lazzeri, P.A., Jager-Gussen, M. and Lorz, H.
Theor. Appl. Genet., 1991 82 47-80
James, M.G., Robertson, D.S. and Myers, A.M.
Plant Cell, 1995 7 417-429
Jolly, C.J., Glenn, G.M. and Rahman, S.
Proc. Natl Acad. Sci., 1996 93 2408-2413.
Kawasaki, T., Mizuno, K., Baba, T. and Shimada, H.
Molec. Gen. Genet., 1993 237 10-16.
Lagudah, E.S., Appels, R. and McNeill, D.
Genome, 1991 34 387-395
Lazzeri, P.A., Brettschneider, R., Luhrs, R. and Lorz, H.
Theor. Appl. Genet., 1991 81 437-444
Maniatis, T., Fritsch, E.F. and Sambrook, J.
Molecular cloning. A Laboratory Manual., New York. Cold Spring Harbor Laboratory, 1982
Marshall,J.; Sidebottom,C.; Debet,M.; Martin,C.; Smith,A.M.; Edwards,A.
The Plant Cell, 1996 8 1121-1135
Martin, C. and Smith, A.
The Plant Cell, 1995 7 971-985.
McElroy, D., Blowers, A.D., Jenes, B., Wu R.

Mol. Gen. Genet., 1991 231 150-160.

Mizuno, K., Kawasaki, T., Shimada, H., Satoh, H., Koyabashi, E., Okumura, S., Arai, Y. and Baba, T.
J.Biol. Chem., 1993 268 19084-19091.

Muller,M.; Knudsen,S.
Plant J, 1993, 4 343-355

Morell, M.K., Blennow, A., Kosar-Hashemi, B. and Samuel, M.S.
Plant Physiol., 1997 113 201-208.

Morell, M.K., Rahman, S., Abrahams, S.L. and Appels, R.
Aust.J.of Plant Physiol., 1995 22 647-660.

Nair, R., Baga, M., Scoles, G.J., Kartha, K. and Chibbar, R.
Plant Science, 1997 1222 153-163

Nakamura,Y.; Kubo,A.; Shimamune,T.; Matsuda,T.; Harada,K.; Satoh,H.
Plant J, 1997, 12 143-153

Nakamura, T., Yanamori, M., Hirano, H., Hidaka, S. and Nagamine, T.
Molecular and General Genetics, 1995 248 253-259

Nakamura, Y., Takeichi, T., Kawaguchi, K. and Yamanouchi, H.
Physiologia Plantarum, 1992 84 329-335.

Nakamura, Y., Umemoto, T. and Sasaki, T.
Planta, 1996 199 209-214

Rahman, S., Kosar-Hashemi, B., Samuel, M., Hill, A., Abbott, D.C., Skerritt, J.H., Preiss, J., Appels, R. and Morell, M.
Aust. J. Plant Physiol., 1995 22 793-803.

Rahman, S., Abrahams, S., Mukai, Y., Abbott, D., Samuel, M., Morell, M. and Appels, R.
Genome, 1997 40 465-474

Repellin, A., Nair, R.B., Baga, M. and Chibbar, R.N.
Plant Gene Register PGR97-094 (1997)

Sambrook, J., Fritsch, E.F. and Maniatis, T.
Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 2nd ed 1989)

Sheen, J., Hwang, S., Niwa, Y., Kobayashi, H., and Galbraith, D.W.
The Plant Journal, 1995 8 777-784

Svensson, B.
Plant Mol. Biol., 1994 25 141-157.
Tanaka, K., Ohnishi, S., Kishimoto, N., Kawasaki, T., Baba, T.
Plant Physiol 1995, 108 677-683
Tingay, S., McElroy, D., Kalla, R., Fieg, S., Wang, M., Thornton, S. and Bretell, R.
The Plant Journal, 1997 11 1369-1376
Wan, Y. and Lemaux, P.G.
Plant Physiology, 1994 104 37-48
Wang, M.B., Upadhyaya, N.M., Brettell, R.I.S., and Waterhouse, P.M.
Journal of Genetics and Breeding, 1997 51 325-334.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION
(B) STREET: Limestone Avenue
(C) CITY: Campbell
(D) STATE: ACT
(E) COUNTRY: AUSTRALIA
(F) POSTAL CODE (ZIP): 2612

(A) NAME: THE AUSTRALIAN NATIONAL UNIVERSITY
(B) STREET: BRIAN LEWIS CRESCENT
(C) CITY: ACTON
(D) STATE: ACT
(E) COUNTRY: AUSTRALIA
(F) POSTAL CODE (ZIP): 2601

(A) NAME: GOODMAN FIELDER LIMITED
(B) STREET: LEVEL 42, GROSVENOR PLACE
(C) CITY: SYDNEY
(D) STATE: NSW
(E) COUNTRY: AUSTRALIA
(F) POSTAL CODE (ZIP): 2000

(A) NAME: GROUPE LIMAGRAIN PACIFIC PTY LIMITED
(B) STREET: LEVEL 31, 1 O'CONNELL STREET
(C) CITY: SYDNEY
(D) STATE: NSW
(E) COUNTRY: AUSTRALIA
(F) POSTAL CODE (ZIP): 2000

(ii) TITLE OF INVENTION: REGULATION OF GENE EXPRESSION IN PLANTS

(iii) NUMBER OF SEQUENCES: 17

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "pcr primer based on the N-terminal sequence of wSBE I 5 ' end at position 168 of SEQ ID NO:5"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE:

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGCACGCGAG AGACTGG 17

(2) INFORMATION FOR SEQ ID NO: 2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "pcr primer in which 5 ' end is at position 1590 of SEQ ID NO:5"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE:

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

TACATTTCCT TGTCCATCA 19

(2) INFORMATION FOR SEQ ID NO: 3:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "pcr primer 5 ' end is at position 1 of SEQ ID NO:5"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE:

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATCACGAGAG CTTGCTCA 18

(2) INFORMATION FOR SEQ ID NO: 4:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "pcr primer 5 ' end is at position 334 of SEQ ID NO:5"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE:

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CGGTACACAG TTGCGTCATT TTC 23

(2) INFORMATION FOR SEQ ID NO: 5:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2687 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE:

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

ATCGACGAAG ATGCTCTGCC TCACCGCCCC CTCCTGCTCG CCATCTCTCC CGCCGCGCCC 60

CTCCCGTCCC GCTGCTGACC GGCCCGGACC GGGGATTTCG GCCAAGAGCA AGTTCTCTGT 120

TCCCGTGTCT GCGCCAAGAG ACTACACCAT GGCAACAGCT GAAGATGGTG TTGGCGACCT 180

TCCGATATAC GATCTGGATC CGAAGTTTGC CGGCTTCAAG GAACACTTCA GTTATAGGAT 240

GAAAAAGTAC CTTGACCAGA AACATTCGAT TGAGAAGCAC GAGGGAGGCC TTGAAGAGTT 300

CTCTAAAGGC TATTTGAAGT TTGGGATCAA CACAGAAAAT GACGCAACTG TGTACCGGGA 360

ATGGGCCCCT GCAGCAATGG ATGCACAACT TATTGGTGAC TTCAACAACT GGAATGGCTC 420
TGGGCACAGG ATGACAAAGG ATAATTATGG TGTTTGGTCA ATCAGGATTT CCCATGTCAA 480
TGGGAAACCT GCCATCCCCC ATAATTCCAA GGTTAAATTT CGATTTCACC GTGGAGATGG 540
ACTATGGGTC GATCGGGTTC CTGCATGGAT TCGTTATGCA ACTTTTGACG CCTCTAAATT 600
TGGAGCTCCA TATGACGGTG TTCACTGGGA TCCACCTTCT GGTGAAAGGT ATGTGTTTAA 660
GCATCCTCGG CCTCGAAAGC CTGACGCTCC ACGTATTTAC GAGGCTCATG TGGGGATGAG 720
TGGTGAGAGG CCTGAAGTAA GCACATACAG AGAATTTGCA GACAATGTGT TACCGCGCAT 780
AAAGGCAAAC AACTACAACA CAGTTCAGCT GATGGCAATC ATGGAACATT CCATATTATG 840
CTTCTTTTGG TACCATGTGA CGAATTTCTT CGCAGTTAGC AGCAGATCAG GAACACCAGA 900
GGACCTCAAA TATCTTGTTG ACAAGGCACA TAGCTTAGGG TTGCGTGTTC TGATGGATGT 960
TGTCCATAGC CATGCGAGCA GTAATATGAC AGATGGTCTA AATGGCTATG ATGTTGGACA 1020
AAACACACAG GAGTCCTATT TCCATACAGG AGAAAGGGGT TATCATAAAC TGTGGGATAG 1080
TCGCCTGTTC AACTATGCCA ATTGGGAGGT CTTACGGTAT CTTCTTTCTA ATCTGAGATA 1140
TTGGATGGAC GAATTCATGT TTGACGGCTT CCGATTTGAT GGAGTAACAT CCATGCTATA 1200
TAATCACCAT GGTATCAATA TGTCATTCGC TGGAAATTAC AAGGAATATT TTGGTTTGGA 1260
TACCGATGTA GATGCAGTTG TTTACATGAT GCTTGCGAAC CATTTAATGC ACAAAATCTT 1320
GCCAGAAGCA ACTGTTGTTG CAGAAGATGT TTCAGGCATG CCAGTGCTTT GTCGGTCAGT 1380
TGATGAAGGT GGAGTAGGGT TTGACTATCG CCTTGCTATG GCTATTCCTG ATAGATGGAT 1440
TGACTACTTG AAGAACAAAG ATGACCTTGA ATGGTCAATG AGTGCAATAG CACATACTCT 1500
GACCAACAGG AGATATACGG AAAAGTGCAT TGCATATGCT GAGAGCCACG ATCAGTCTAT 1560
TGTTGGCGAC AAGACTATGG CATTTCTCTT GATGGACAAG GAAATGTATA CTGGCATGTC 1620
AGACTTGCAG CCTGCTTCAC CTACAATTGA TCGTGGAATT GCACTTCAAA AGATGATTCA 1680
CTTCATCACC ATGGCCCTTG GAGGTGATGG CTACTTGAAT TTTATGGGTA ATGAGTTTGG 1740
CCACCCAGAA TGGATTGACT TTCCAAGAGA AGGCAACAAC TGGAGTTATG ATAAATGCAG 1800
ACGCCAGTGG AGCCTCTCAG ACATTGATCA CCTACGATAC AAGTACATGA ACGCATTTGA 1860
TCAAGCAATG AATGCGCTCG ACGACAAGTT TTCCTTCCTA TCGTCATCAA AGCAGATTGT 1920
CAGCGACATG AATGAGGAAA AGAAGATTAT TGTATTTGAA CGTGGAGATC TGGTCTTCGT 1980
CTTCAATTTT CATCCCAGTA AAACTTATGA TGGTTACAAA GTCGGATGTG ATTTGCCTGG 2040
GAAGTACAAG GTAGCTCTGG ACTCCGATGC TCTGATGTTT GGTGGACATG GAAGAGTGGC 2100
CCAGTACAAC GATCACTTCA CGTCACCTGA AGGAGTACCA GGAGTACCTG AAACAAACTT 2160
CAACAACCGC CCTAATTCAT TCAAAGTCCT GTCTCCACCC CGCACTTGTG TGGCTTACTA 2220
TCGCGTCGAG GAAAAAGCGG AAAAGCCTAA GGATGAAGGA GCTGCTTCTT GGGGCAAAGC 2280
TGCTCCTGGG TACATCGATG TTGAAGCCAC TCGTGTCAAA GACGCAGCAG ATGGTGAGGC 2340

GACTTCTGGT TCCAAAAAGG CGTCTACAGG AGGTGACTCC AGCAAGAAGG GAATTAACTT 2400

TGTCTTCGGG TCACCTGACA AAGATAACAA ATAAGCACCA TATCAACGCT TGATCAGAAC 2460

CGTGTACCGA CGTCCTTGTA ATATTCCTGC TATTGCTAGT AGTAGCAATA CTGTCAAACT 2520

GTGCAGACTT GAGATTCTGG CTTGGACTTT GCTGAGGTTA CCTACTATAT AGAAAGATAA 2580

ATAAGAGGTG ATGGTGCGGG TCGAGTCCGG CTATATGTGC CAAATATGCG CCATCCCGAG 2640

TCCTCTGTCA TAAAGGAAGT TTCGGGCTTT CAGCCCAGAA TAAAAAA 2687

(2) INFORMATION FOR SEQ ID NO: 6:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 807 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE:

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(ix) FEATURE:
(A) NAME/KEY: Protein
(B) LOCATION: 1..807
(D) OTHER INFORMATION:/label= sbeI
/note= "deduced amino acid sequence from SEQ ID NO:5"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Leu Cys Leu Thr Ala Pro Ser Cys Ser Pro Ser Leu Pro Pro Arg
1               5                   10                  15

Pro Ser Arg Pro Ala Ala Asp Arg Pro Gly Pro Gly Ile Ser Ala Lys
            20                  25                  30

Ser Lys Phe Ser Val Pro Val Ser Ala Pro Arg Asp Tyr Thr Met Ala
        35                  40                  45

Thr Ala Glu Asp Gly Val Gly Asp Leu Pro Ile Tyr Asp Leu Asp Pro
    50                  55                  60

Lys Phe Ala Gly Phe Lys Glu His Phe Ser Tyr Arg Met Lys Lys Tyr
65                  70                  75                  80

Leu Asp Gln Lys His Ser Ile Glu Lys His Glu Gly Gly Leu Glu Glu
                85                  90                  95

Phe Ser Lys Gly Tyr Leu Lys Phe Gly Ile Asn Thr Glu Asn Asp Ala
            100                 105                 110

Thr Val Tyr Arg Glu Trp Ala Pro Ala Ala Met Asp Ala Gln Leu Ile
        115                 120                 125
```

```
Gly Asp Phe Asn Asn Trp Asn Gly Ser Gly His Arg Met Thr Lys Asp
    130                 135                 140

Asn Tyr Gly Val Trp Ser Ile Arg Ile Ser His Val Asn Gly Lys Pro
145                 150                 155                 160

Ala Ile Pro His Asn Ser Lys Val Lys Phe Arg Phe His Arg Gly Asp
                165                 170                 175

Gly Leu Trp Val Asp Arg Val Pro Ala Trp Ile Arg Tyr Ala Thr Phe
            180                 185                 190

Asp Ala Ser Lys Phe Gly Ala Pro Tyr Asp Gly Val His Trp Asp Pro
        195                 200                 205

Pro Ser Gly Glu Arg Tyr Val Phe Lys His Pro Arg Pro Arg Lys Pro
    210                 215                 220

Asp Ala Pro Arg Ile Tyr Glu Ala His Val Gly Met Ser Gly Glu Arg
225                 230                 235                 240

Pro Glu Val Ser Thr Tyr Arg Glu Phe Ala Asp Asn Val Leu Pro Arg
                245                 250                 255

Ile Lys Ala Asn Asn Tyr Asn Thr Val Gln Leu Met Ala Ile Met Glu
            260                 265                 270

His Ser Ile Leu Cys Phe Phe Trp Tyr His Val Thr Asn Phe Phe Ala
        275                 280                 285

Val Ser Ser Arg Ser Gly Thr Pro Glu Asp Leu Lys Tyr Leu Val Asp
    290                 295                 300

Lys Ala His Ser Leu Gly Leu Arg Val Leu Met Asp Val Val His Ser
305                 310                 315                 320

His Ala Ser Ser Asn Met Thr Asp Gly Leu Asn Gly Tyr Asp Val Gly
                325                 330                 335

Gln Asn Thr Gln Glu Ser Tyr Phe His Thr Gly Glu Arg Gly Tyr His
            340                 345                 350

Lys Leu Trp Asp Ser Arg Leu Phe Asn Tyr Ala Asn Trp Glu Val Leu
        355                 360                 365

Arg Tyr Leu Leu Ser Asn Leu Arg Tyr Trp Met Asp Glu Phe Met Phe
    370                 375                 380

Asp Gly Phe Arg Phe Asp Gly Val Thr Ser Met Leu Tyr Asn His His
385                 390                 395                 400

Gly Ile Asn Met Ser Phe Ala Gly Asn Tyr Lys Glu Tyr Phe Gly Leu
                405                 410                 415

Asp Thr Asp Val Asp Ala Val Val Tyr Met Met Leu Ala Asn His Leu
            420                 425                 430

Met His Lys Ile Leu Pro Glu Ala Thr Val Val Ala Glu Asp Val Ser
        435                 440                 445

Gly Met Pro Val Leu Cys Arg Ser Val Asp Glu Gly Gly Val Gly Phe
    450                 455                 460
```

Asp Tyr Arg Leu Ala Met Ala Ile Pro Asp Arg Trp Ile Asp Tyr Leu
465 470 475 480

Lys Asn Lys Asp Asp Leu Glu Trp Ser Met Ser Ala Ile Ala His Thr
485 490 495

Leu Thr Asn Arg Arg Tyr Thr Glu Lys Cys Ile Ala Tyr Ala Glu Ser
500 505 510

His Asp Gln Ser Ile Val Gly Asp Lys Thr Met Ala Phe Leu Leu Met
515 520 525

Asp Lys Glu Met Tyr Thr Gly Met Ser Asp Leu Gln Pro Ala Ser Pro
530 535 540

Thr Ile Asp Arg Gly Ile Ala Leu Gln Lys Met Ile His Phe Ile Thr
545 550 555 560

Met Ala Leu Gly Gly Asp Gly Tyr Leu Asn Phe Met Gly Asn Glu Phe
565 570 575

Gly His Pro Glu Trp Ile Asp Phe Pro Arg Glu Gly Asn Asn Trp Ser
580 585 590

Tyr Asp Lys Cys Arg Arg Gln Trp Ser Leu Ser Asp Ile Asp His Leu
595 600 605

Arg Tyr Lys Tyr Met Asn Ala Phe Asp Gln Ala Met Asn Ala Leu Asp
610 615 620

Asp Lys Phe Ser Phe Leu Ser Ser Ser Lys Gln Ile Val Ser Asp Met
625 630 635 640

Asn Glu Glu Lys Lys Ile Ile Val Phe Glu Arg Gly Asp Leu Val Phe
645 650 655

Val Phe Asn Phe His Pro Ser Lys Thr Tyr Asp Gly Tyr Lys Val Gly
660 665 670

Cys Asp Leu Pro Gly Lys Tyr Lys Val Ala Leu Asp Ser Asp Ala Leu
675 680 685

Met Phe Gly Gly His Gly Arg Val Ala Gln Tyr Asn Asp His Phe Thr
690 695 700

Ser Pro Glu Gly Val Pro Gly Val Pro Glu Thr Asn Phe Asn Asn Arg
705 710 715 720

Pro Asn Ser Phe Lys Val Leu Ser Pro Pro Arg Thr Cys Val Ala Tyr
725 730 735

Tyr Arg Val Glu Glu Lys Ala Glu Lys Pro Lys Asp Glu Gly Ala Ala
740 745 750

Ser Trp Gly Lys Ala Ala Pro Gly Tyr Ile Asp Val Glu Ala Thr Arg
755 760 765

Val Lys Asp Ala Ala Asp Gly Glu Ala Thr Ser Gly Ser Lys Lys Ala
770 775 780

Ser Thr Gly Gly Asp Ser Ser Lys Lys Gly Ile Asn Phe Val Phe Gly
785 790 795 800

Ser Pro Asp Lys Asp Asn Lys
805

(2) INFORMATION FOR SEQ ID NO: 7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 319 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE:

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(ix) FEATURE:
(A) NAME/KEY: misc_signal
(B) LOCATION:1..319
(D) OTHER INFORMATION:/function= "3' untranslated region of wSBE I-D4 cDNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
GCGACTTCTG GTTCCAAAAA GGCGTCTACA GGAGGTGACT CCAGCAAGAA GGGAATTAAC   60
TTTGTCTTCG GGTCACCTGA CAAAGATAAC AAATAAGCAC CATATCAACG CTTGATCAGA  120
ACCGTGTACC GACGTCCTTG TAATATTCCT GCTATTGCTA GTAGTAGCAA TACTGTCAAA  180
CTGTGCAGAC TTGAGATTCT GGCTTGGACT TTGCTGAGGT TACCTACTAT ATAGAAAGAT  240
AAATAAGAGG TGATGGTGCG GGTCGAGTCC GGCTATATGT GCCAAATATG CGCCATCCCG  300
AGTCCTCTGT CATAAAGGA                                               319
```

(2) INFORMATION FOR SEQ ID NO: 8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4890 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE:

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(ix) FEATURE:

(A) NAME/KEY: promoter
(B) LOCATION:1..4890
(D) OTHER INFORMATION:/function= "promoter containing sequence of SBE I"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
GGGTGGCGGG TCGGGCGGCA AGGCGCGGGG CGGCGGGGCG GCCGGGGCGG CGCGGCGGCG   60
CGGGCGGCAG CGGCGGCTAG GGTTTCGCGG CGGCGGCGAC TTGGGCTGAG GCGGGGCACG  120
GGCTGCGGCT TTAAAGGCCG GCCAGGCTGA GGTGTCCGGG TCGGACACGG CCCGTAAGGC  180
GGTTGACTTT AAAAAATAAT AATTCGGACA TGCAAAAAAG TAAGAAAAGA AATAATAAAC  240
GGACTCCAAA AATCCCGAAG TAAATTTTTC CCCATTCTTA AAAATAAGCC GGACAAGATG  300
AACATTTATT TGGGCCTAAA ATGCAATTTT GAAAAATGCG TATTTTTCCT AATTCGGAAT  360
AAAATCAAAT AAAATCCAAA TAAAATCAAA TATTTGTTTT TAATATTTTT CCTCCAATAT  420
TTCATTATTT GTGAAGAAGT CATTTTATCC CATCTCATAT ATTTTGATAT GAAATATTTT  480
CGGAGAGAAA AATAATTAAA ACAAATGATC CTATTTTCAA AATTTGAGAA AACCCAAATA  540
TGAAAATAAC GAAATCCCCA ACTCTCTCCG TGGGTCCTTG AGTTGCGTGA AATTTCTAGG  600
ATCACAAATC AAAATGCAAT AAAATATGAT ATGCATGATG ATCTAATGTA TAACATTCCA  660
ATTGAAAATT TGGGATGTTA CATATAACTC AAATTCTATA ATTATGAACA CAGAAATATT  720
AATGTAGAAC TCTATTTTGT TTTGAAATTG TATTATTTTT TAGAATTAGT CTAGAGCATT  780
TCGTGAACTT GAATCAAACC TTTAAATAAA ACAAAGCATA AAAATGACAA ATTCACATAT  840
GAAATAACTT GTGTTACATA GATTTATTAC AATAGCGTTG TATGTGTGTA TGTGTGCGTG  900
AGTGCCTATG GTAATATCAA TAAATATCTT GATAGATGTT TCTACAATTC ACGGGTCTAA  960
CTAGTAATGC AATGCAATGC ATGCTAAAAG AATAGAACCT TAGTTTCATT TAACTAACAA 1020
TTTTCAAATG TATGAGTTGC CAACAAGTGG CATACTTGGC ACTGTTTGTT TGTTCATTTT 1080
ATGGAAAGTT CTTCTCTTTT TACATGGTTT AGATTCCAGC ATGTAGCCAC AAAATATGAT 1140
TGTCAAAAGA TAATACCTCA TAATACAATT CCACTAAAGT CACCTAGCCC AAGTGACCGA 1200
CCTGATCCTG AAATAAAATC AGAAGATTTG GTGTCATCAT CATGACAACA AATTATTAGG 1260
CGGTAGATCT TGTGGTAGTA CTCATGATGT AAAATTATCA AGAGGGAGAG AATGTATGGA 1320
GATTTATGTG AAGTACATCG TACACCAGAC ATAGTTGACA CATCGATTTT TTAAGATACA 1380
TTTGGACGCG CCTTGTGGGA GTGTAAAGTA CTACCATGTA TTAGAAGAGG TGAAATGAGA 1440
AATGCCATAG CTAGCAAGTA GGCCTAGTTA AGGAAATTCT TCCTTAGATC CCCTTCTCCC 1500
GAAGAGTGAA GTGCTTCAAC TAAAGGTTAG ACCCACTTAA AAAATGTCAC TTTGAATCTT 1560
TGCTTCCCTT GTCGTAATCC TGTGCATTTG TAGGTCCCTC GGATCTGAGC CCTTTCTCCA 1620
AGCCCTTCAT TGGATTCCCC TGGATGTCTT TTTGTTACAT TTTATTGAAG TGAGAGTGAA 1680
TTATTATATG CCCATAGGAG GTGGGATATA AAGGCTGTTG GTATTCTGCA CCATACATGC 1740
```

TAGAGTAGGG AGGAGAGGCT GGTGCATGAT ACATGGTGGA CTAGCCCATA TATTTACCCC 1800
TCCCCCACCC ACTAACAAGT TTTTTTTATT AGGTCTTCAT CCTCTGATTT GTTTTTCTGT 1860
TAGCCCATTC TTCATCATGG ACTTATTAAT CATGATTAGT TTCTTGGATT TTTGTTTACT 1920
TGACTTGAAT TTGACAATGT GCCTCATATA TGGCATGTGG GACTGATAGG AAGATATATT 1980
CTCACAACAT TAACTTAAAA AGGATTATTT TTTTGGTGCA GTCGTAAAGA AAACTACTTT 2040
CTTTTATGCT AAAAGTTATT CAAACATAGA TTTATAAACA AAGGATATCA CCATGCATGA 2100
CCATGCGCTC TCTCATGTTT ACTCTAGAAA CCATATATCT CTTTGTTGCA AAATATTTAA 2160
TCTATCCTCC TTGTTTCTGG GAATGAGTCG GGGAAGGTAA TCTTAGGGAA GGTTAAAGTG 2220
AGGCAAGTAA GAGCAACTCT AGCAGAGTCG CGATATGCCC AATCGCCATA ATGCCAATAT 2280
GGCATTTTTG GCCCAAAATG GCACTTCAGA AGAGTCACCA TATCCCTTCG GATAGCCATA 2340
ATTTAGGGAG CTCGCTCCAC AAACAAGCTT CGAGCCTCCA AATATGGAGG CCATGGATTC 2400
GTTGTTTGGC ACTCACTCCA TATCCAACCG CAAGCGCATG CATGAGGGAA GTTTTAGCTT 2460
CTTCCTCCTT GCGCCAACGC CGGGATTTTA CACAGCGCAT TACAGGTACA TGAACCAGCA 2520
TGCACAGATA ATCACCGACG AGTGGGGTGA CAAGAAGGAT AAGCACCCTC CCATTAGTGG 2580
TGCGCCCACT CCCCTCAAAT TCATGAGGCA GCCATTTGGA TGGTCATCGC GTGGCATAAG 2640
CTCCGACTAT AAAATCTCAA CGGCATCACC AAAACCATAG CTGCCGCCTC CCCCTTCCTC 2700
GGCATCACCT CCCCAAGACA TCTCCTCCCC TCTATGCCAC AATGTCATCA TTATGGAGAG 2760
ACACAACTAC TGGTAAACCG CATACCCAAT CATGGTTTAC CGGCAGTGCG AACCCCACCT 2820
TCCTCCCACG ATGGTAGGAT ATTCTCCTCC TAGAATGGCG CGTGTGGCGC TTCCTCCTCC 2880
CGAGGCTGAT ATGTCGGCTC CCATGATGGC GTGCATCATT GATTTGGCGC TTCGGGTCCA 2940
TCATACATGT TAACGAGGTC ATCCCCATTG ATGTCGTTGG TCCCCTTGCC CCCCAGTCGG 3000
ATCCTGAGGA CCCGTTCGAT GTCGCAATGC GACTCTCCAA ACTCAAAGCT CACAATGAGG 3060
AGTACGTCCT CTAGGAGTTC CGCCCCGCAA CCATCTATAA GGAGGAGCAA CGATAGCTCT 3120
CCCCTACGCC TTCCTCGACG ATCTCTCTTA GGAGGACAAC GGCTAGACGA CGGCGGCGGC 3180
GGCGAAGGTA CTGCAGGTAG TAGAACATAG CAATGTCGAA TGGCGACATT GCATATTTTG 3240
AAAATGTCGC TCAACGACTT TTGAAGTCGC AAATAAAATG TAGTGTGACT ACTTTTGGCC 3300
AGCAATATAA GTTTATCACA TTTGATAATG ATTTGAACCG GTGTGGTTCA ACTAAATGTA 3360
CCATAAATTG AACATACAAA TTTTTAGCAA ATGAAAAAAG AAACAAGTAA GACCACAAAT 3420
ATGAAAGCCG CATATCGCGA CTATGTGTTT GAGCCGCAGC TGCCAAGTAC ATATGAAGCG 3480
TACTCCATAT GACATACGAC AACCATACAT ATGAAGACTC TACTAGAGTT CTCTAAGGCC 3540
GCTTTTAGCG CCTTTCGTGC AGTGGTGCCC ATAGGGAGTG AGGGTAGTTG GACTGTTCGT 3600
TTCCCCTTTT TTCATTTCTT TGAAATCTAT TTTATTTTTT TTCTCTTTTG TAGGTTTCCC 3660
AAATTTATAT ACCATTTTTC TGTTTCTCGC TATTTTTTGT TGTTATATTC TAGTTTCATA 3720
TTTTTCTATT ATTAATTTGT GTCTCTTATG AGAAGTCCAG ACTTGCATAT GGAGGTGCAC 3780

ACACAAACAT ATAAAGTATA AATACTAACT TGAGAAGTAT GTTTGCGTGG TCAAAAAAAC 3840
ATCATCAAAA CCTGCCAATA TGAGATATAG TTTTGAATAT ATCAATATGA GCAACGCAAC 3900
CATTTAAAAT GTGAACAATT GTTTTTTTAG AAAAAATATA AGAAATAACT CCAACCCAGC 3960
CAAACCACAT GCTATACACT TGCTCCATAT GAAACCATGT TTGCTATTGG GCAGTTGCCT 4020
GAAACCGAAA GTAATGTTAG CCGTTTTTCT ATTCAAAGAA GAAGGAGAGT CGAGGTGACG 4080
CGATGCTTAG ACGTGAGATG GGGATGACCA CAACGTCCCT ACAGAGACCT CACCGGAGAT 4140
GGGGACATTG CAGTTGACAC GAGAGCGGTG AGGGGCTGCG ATGCGTGTGC GGCAACATGT 4200
GGCGAGGCGG ACGTCGGGCT GGCAGGTAGG GGGGAGGGGG AAGGACCGGG GGAGGAAGAA 4260
GAGGAGTAGC CTGCAAAACA TGGTACACCA GTTTTCTGCC CTACGAAAAC CTCATTTCAT 4320
TCCCCCACCC TGACAAGCAA CAACCAACCA TCGCAGTCCC ACATGTCCCT CTGGTCTTTG 4380
CAAAAAGTAA TTGTTCTTGC TGGACAGCGC AAAGAGTAAA CTTTTGTTAG TTTTCATTTC 4440
TAGAAAAAGC AATCCTTTTA TAGTTCTTTT GTGAAAGTAA TGCTTTTATA GTGATTGGGA 4500
TGTTCTTTTA GAGCAAATAT CTTCTTTTTT TTTTAGGGAA AAGAGCAAAT ATCTTCCACT 4560
TTTCACAAAA CTGACGAAGG CTGAAAGTGG CGAGACAGTG AGGGCCCATA GCTTTCGTCC 4620
GGCCCAGCGG CGCACGACCG TCCACGTGCA CCCCGGCCCT CCCGGGCCCG CAGATCCGTT 4680
CTCCCTCGCC CCCGTTTCCC CCTCCCTCCC TCTCGTTGCT TCCACTCCAC TGTTCTCCTC 4740
TTCCTGTCCA AAGCGGCCAC GGACCGGAAA AAAATCACGC CTTTCCGTTG GGTCTCCGGC 4800
GCCACACTCC TCCTCCGGCC GATATAAAGC GCGCGGGGCC ACGGGCCCGG CGCAAAATGG 4860
GATTCCCGTC CGCCGCCATG GAGGAAGATG 4890

(2) INFORMATION FOR SEQ ID NO: 9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6228 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE:

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1
(D) OTHER INFORMATION:/product= "coding region of wSBE I-D4 gene"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

ACGGGCCCGG CGCAAAATGG GATTCCCGTC CGCCGCCATC GACGAAGATG CTCTGCCTCA 60
CCGCCCCCTC CTGCTCGCCA TCTCTCCCGC CGCGCCCCTC CCGTCCCGCT GCTGACCGGC 120
CCGGACCGGG GATCTCGGTG AGTCAGTCGG GATCTTCATT TCTTTTCTTT TCTTTCGTTT 180
CCGGCTCCGT TCTGCCGGGG TTTCCCTGAT GCGATGCCGC GCGCGCGCAG GGCGGCGGCA 240
ATGTGCGGCT GAGCGCGGTG CCCGCGCCCT CTTCGCTCCG CTGGTCGTGG CCGCGGAAGG 300
TGAGCCCTCT CCCCTGTCTA CCCAGATTTG CGACCGTGAT CCCCTGTTGT CGCCGGGCAA 360
ACGGAATCTG ATCCACGGTG GTTATTGGAA ATAGTATATA CTACTAATAA ACTTGAGGCT 420
GGGATTCGTC CACTGAGGAA CAAGTGGATG CGATTTCGAT TGGATTTCTC TGCTTTATGC 480
GATCCGTACG CAGAATATCC CTCCTGCAGT GTCTCAACCG TATTACTGGA TGTACAACCC 540
AAATGTGTAT AATCTGTGCT GAATGTATCA ACCAATAATT GCTGCATTGT GAAAACATAA 600
TCCTGTGTTG TGTCTCTACT ACTTGTTCAG TCCTGATCTG CCGCTTATCC TAACTTTTGT 660
TCATTTATGG AAGGCCAAGA GCAAGTTCTC TGTTCCCGTG TCTGCGCCAA GAGACTACAC 720
CATGGCAACA GCTGAAGATG GTGTTGGCGA CCTTCCGATA TACGATCTGG ATCCGAAGTT 780
TGCCGGCTTC AAGGAACACT TCAGTTATAG GATGAAAAAG TACCTTGACC AGAAACATTC 840
GATTGAGAAG CACGAGGGAG GCCTTGAAGA GTTCTCTAAA GGTTAGCTTT TGTTTCATGT 900
GTTTGAAACA ATAGTTACAT CTTGTGGCGT CCGCAGCACA AAAGACATAA TGCGACTCTG 960
TTTTGTAGGC TATTTGAAGT TTGGGATCAA CACAGAAAAT GACGCAACTG TGTACCGGGA 1020
ATGGGCCCCT GCAGCAATGT AAGTTCTAGT GTTGTCACGC AACTAATTGC AATGGTCGTT 1080
GGTTAACTTA TGAAGTGCTG ATGAAACTGT CTTAAGAGTT TATGGCTTGT CTTTTCTGAT 1140
TCTAGCTAGT AAAGAGTAGA TAAATATGAA ATATGTTTTC CCTTTTCTAG TTATGGTCAT 1200
GGTTGGCTGG TATTCATTTC TTTTATGGCA ATACTTGCTT CTAACTATCT TTAGTAGATT 1260
CATGTATTTA CTTGTGAGTC ATTACTTTAT GGGTGTAGGG ATGCACAACT TATTGGTGAC 1320
TTCAACAACT GGAATGGCTC TGGGCACAGG ATGACAAAGG ATAATTATGG TGTTTGGTCA 1380
ATCAGGATTT CCCATGTCAA TGGGAAACCT GCCATCCCCC ATAATTCCAA GGTTAAATTT 1440
CGATTTCACC GTGGAGATGG ACTATGGGTC GATCGGGTTC CTGCATGGAT TCGTTATGCA 1500
ACTTTTGATG CCTCTAAATT TGGAGCTCCA TATGACGGTG TTCACTGGGA TCCACCTTCT 1560
GGTGAAAGGT CTACTTTTAG TGGCTCGAGA GCAAGAAATC TAAGTAAAAC CCACACAATT 1620
AACTTACATT AATGTGGAGA CATGATACTT TTATTGCTCG TTTTGCAGGT ATGTGTTTAA 1680
GCATCCTCGG CCTCGAAAGC CTGACGCTCC ACGTATTTAC GAGGCTCATG TGGGGATGAG 1740
TGGTGAAAAG CCTGAAGTAA GCACATACAG AGAATTTGCA GACAATGTGT TACCGCGCAT 1800
AAAGGCAAAC AACTACAACA CAGTTCAGCT GATGGCAATC ATGGAACATT CATATTATGC 1860
TTCTTTTGGG TACCATGTGA CGAATTTCTT CGCAGTTAGC AGCAGATCAG AACGCCAGAG 1920
ACCTCAATAT CTTGTTGACA AGGCACATAG TTTACGGTTG CGTGTTCTGA TGGATGTTGT 1980
CCATAGCCAT GCGAGCAGTA ATAAGACAGA TGGTCTTAAT GGCTATGATG TTGGGCAAAA 2040

CACACAGGAG TCCTATTTCC ACACAGGAGA AAGGGGCTAT CATAAACTGT GGGATAGCCG 2100
CCTGTTCAAC TATGCCAATT GGGAGTCTTA CGATTTCTTC TTTCTAATCT GAGATATTGG 2160
ATGGACGAAT TCATGTTTGA TGGCTTCCGA TTTGATGGGG TAACATCCAT GCTATATAAT 2220
CACCATGGTA TCAATATGTC ATTCGCTGGA AGTTACAAGG AATATTTTGG TTTGGATACT 2280
GATGTAGATG CAGTTGTTTA CCTGATGCTT GCGAACCATT TAATGCACAA ACTCTTGCCA 2340
GAAGCAACTG TTGTTGCAGA AGATGTTTCA GGCATGCCAG TGCTTTGTCG GTCAGTTGAT 2400
GAAGGTGGAG TAGGGTTTGA CTATCGCCTG GCTATGGCTA TTCCTGATAG ATGGATCGAC 2460
TACTTGAAGA ACAAAGATGA CCTTGAATGG TCAATGAGTG GAATAGCACA TACTCTGACC 2520
AACAGGAGAT ATACGGAAAA GTGCATTGCA TATGCTGAGA GCCATGATCA GGTATGTTTT 2580
CCCTCCTTTG TCGCTGTGCG TGAGTATGTG TTCTTTTTTT ATGGGGCACT GGTCTAAGAA 2640
CATACAGTTC AAAGGTGAGA CACTTTCTTT GCCTGGTAGA CAAATTTGAG AAATAAACAT 2700
TTCGCTTGAT GACTTTTAGT TGCTTCACAA GTTCGAATTA AGTTAGTTAT ATTCTGATAA 2760
CTAGTGATAG TACCCACTAA CCAGCTATTA CGGACCATGT AAGAATGTCC GAAGACTGCA 2820
GTTATATATC GTTGACTTTG TGTTCATCTA TTGAAACAAC TTAGTAGTTA ACTTTCACGC 2880
AAATTTTCAG TCTATTGTTG GCGACAAGAC TATGGCATTT CTCTTGATGG ACAAGGAAAT 2940
GTATACTGGC ATGTCAGACT TGCAGCCTGC TTCGCCTACA ATTGATCGTG GAATTGCACT 3000
TCAAAAGGTT CGATTCGTTT TAAGTATTCC TGAATTTGAT GTTCTAGTTC CAGACGAGTA 3060
TTGTAATGTT CGTTGTTACT CAGAGTTCTG CTTAGTCCTT GAAGATAATG TATTCCAGTC 3120
CCTTTTGGTA CATTTGGCTT ATTTTGTTAC AAATATTTCA GATGATTCAC TTCATCACCA 3180
TGGCCCTTGG AGGTGATGGC TACTTGAATT TTATGGGTAA TGAGGTAATA TCTGGTTATC 3240
TGTCAAAACT TATTTCTGAT CAATATGTTT CGGGATTCCC TCGAAAAAAA TCCTTTGGGC 3300
AGGGCGAAAA GTTTAAACAT CTGTTTTCTA TGATAGCCAA GTACTCCCCA GCTATTTCCA 3360
TGTTATCACG TATCATTTAG CTGTGCCGGT AGTTAATCTT TATTCTAATT CATTGTTGTT 3420
TTTTAGCGTG GCAGTCTATT GTTGGATCCT CTTATTCCAA TTACATATAT GCCGACATCA 3480
CACACTTATG AATATTCCCT GTTTAAAAGA TTTTTATTTT ATACCAATGT TTCTCCGTAA 3540
ATGATGCAAA CATGATAGAG ATGTTAGCAT GTCTTTCTTA ACCTACTCAT GTTTTACATA 3600
TCACGACAAG CTTCTTGCAG AAAATCAGCA GTATATGGCA AATTGCTGCA ACCTGACAAC 3660
GTTTATATCT GTTTTCTAAC TCATACTGAC GGTGCAATTT CCTTTTAGTT TGGCCACCCA 3720
GAATGGATTG ACTTTCCAGA AGAAGGCAAC AACTGGAGTT ATGATAAATG CAGACGCCAG 3780
TGGAGCCTCG CAGACATTGA TCACCTACGA TACAAGGTTA TGCCTATGTA TATTTTTACA 3840
GTTTCTGGTC TGGTAGCTCT CTTGGGATCT TGACCTCACT TAGTTCCTTC ATCTCTGACT 3900
GTAGCTTATT TACACTGTGT TCCAACTTCT GTCTTGTGGA TAAATTCTCC CTTCTAACGT 3960
TTCATATTAA GCCTTTCAAA CTAAACTAAA TTGCTGATCT ACTACTAGTT GCTCAGTACG 4020

ATGACCAAAT CTTGCCTGTG GTAACCTAGT AATTTTCTTG ATTCTTACAC ATTAGTGATA 4080
TGCAGTGCAT ACATTATCCA TATAAATTGA CATTGCAATT TCCCAAATAT TATTTGAAGG 4140
CTGTGTTCTT TTGTTAACAG GAAGTTATTT TCTCTGCATC TGATAAATAA TAATAGCCTT 4200
TCACGATTTT TCTCATATTT TATCCAACTT TTCTGCATTC AAGCATTTTT TGTTTCTCGC 4260
CTAACATATA TAATTTGAAC AGTACATGAA CGCATTTGAT CAAGCAATGA ATGCGCTCGA 4320
CGACAAATTT TCCTTCCTAT CATCATCAAA GCAGATTGTC AGCGACATGA ATGAGGAAAA 4380
GAAGTAGTTA ACTATACAAT GTTTAGTCAG GGCAGCTGTT GCATCATTTG ATTCACTCCT 4440
ACTCTTAAGA ATAGCAACTC TGACTTGTGC GTTTTATGTT ACCAAATAAG TTGAAACCGT 4500
ATCTGTTTGA TATGAACCAT TGTTGTCTCA AAATGGGCTA TGGACTCAAT CCAACTTCCT 4560
TTCCAGATTA TTGTATTTGA ACGTGGAATC TGGTCTTCGT CTTCAATTTT CATCCCAGTA 4620
AAACTTATGA TGGGTAACTG ATCTCTTGCA AGCTTTGCCT TTCAATATTT CTTCTGCTTA 4680
ATGACTAATG TGCTTAATCT CGTTTCCACT TTTAAAACAC GCAGTTACAA AGTCGGATGT 4740
GACTTGCCTG GGAAGTACAA GGTAGCTCTG GACTCTGATG CTCTGATGTT TGGTGGACAT 4800
GGAAGAGTAA GCAATGTTAA TGATGTTCAA GATCTGTTTT GCAACACTAT GTTCTTCTAT 4860
AGAAGGGGCC ATCAAGGCTG CATCAGATAA TCTTATTTGC AGTGTTGATC TGTGCTGCAT 4920
CGCAGGTGGC CCATGACAAC GATCACTTTA CGTCACCTGA AGGAGTACCA GGAGTACCTG 4980
AAACAAACTT CAACAACCGC CCTAACTCAT TCAAAATCCT GTCTCCATCC CGCACTTGTG 5040
TGGTAATGCT AATTACTAGG AGGATTTAGT AACAATAAAT AAATAACAGC AAAAGATATC 5100
TGCAGTACGA TCTCACAAAA TGCTCTCTTG CCAGGCTTAC TATCGCGTCG AGGAGAAAGC 5160
GGAAAAGCCC AAGGATGAAG GAGCTGCTTT CTTGGGGGAA ACTGCTCTCG GGTACATCGA 5220
TGTTGAAGCC ACTGGCGTCA AAGACGCAGC AGATGGTGAG GCGACTTCTG GTTCCGAAAA 5280
GGCGTCTACA GGAGGTGACT CCAGCAAGAA GGGAATTAAC TTTGTCTTTC TGTCACCCGA 5340
CAAAGACAAC AAATAAGCAC CATATCAACG CTTGATCAGG ACCGTGTGCC GACGTCCTTG 5400
TAATACTCCT GCTATTGCTA GTAGTAGCAA TACTGTCAAA CTGTGCAGAC TTGAAATTCT 5460
GGCTTGGACT TTGCTGAGGT TACCTACTAT ATAGAAAGAT AAATAAGCGG TGATGGTGCG 5520
GGTCGAGTCC AGCTATATGT GCCAAATATG CGCCATCCCG AGTCCTCTGT CATAAAGAAA 5580
GTTTCGGGCT TCCATCCCAG AATAAAAACA GTTGTCTGTT TGCAATTTCT TTTTGTCTTG 5640
CATAGTTACA TGATAATTGA TGCATATTGC TATAAGCCTG GATTGCATCT TCTTTTGCTA 5700
ATAACTGCAG GGCCAAGAAA GCCTAGATTG TATCTTTTTT TGCTAATAAC TGCAGTGCTG 5760
GGGAAGCTTC AGTCCTTGTT TCCGTTCTCG AGACAAGGCG TCATGTTTGG CGCACAAAGG 5820
TAAGCCATCA TCTTATCAAG TCCCAAAATT CTCTGGTTGA AAGAAACCAT CACTAACTTG 5880
TTCCAGGTGT TGGTTCCTCC ACAACCAAAA GGCGACCATC GTCGTCATCA TCGCTCACAG 5940
CACTGACCAT CGAAGCCACG GTGGGCATGA AATGCGCATC GCCCAAGACT TGGGACCGTT 6000
TCAAAATATC ACAAACTGCC ATGGCATCTT CTGCCAAAGG CTGCACTGCA CCTTTGGCAT 6060

GAACAGAAGC AACAGGGGCT TGGAACTGAA CGCCGAAAAT AAAGTCAAAC CGGCTGGGCC 6120

GGATTGAAAG GGGAAACGCC AAAATCCACT TAATTTGAAT GGAAGGAGGA ATGGTTCTTG 6180

CTGGTTTCAA CTCTGCAGGC TTCCCTCTGA ATTTCACACG GAGCCATT 6228

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11463 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE:

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..11463
(D) OTHER INFORMATION:/product= "complete sequence of the starch branching enzyme II gene"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

AGAAACACCT CCATTTTAGA TTTTTTTTTT GTTCTTTTCG GACGGTGGGT CGTGGAGAGA 60

TTAGCGTCTA GTTTTCTTAA AAGAACAGGC CATTTAGGCC CTGCTTTACA AAAGGCTCAA 120

CCAGTCCAAA ACGTCTGCTA GGATCACCAG CTGCAAAGTT AAGCGCGAGA CCACCAAAAC 180

AGGCGCATTC GAACTGGACA GACGCTCACG CAGGAGCCCA GCACCACAGG CTTGAGCCTG 240

ACAGCGGACG TGAGTGCGTG ACACATGGGG TCATCTATGG GCGTCGGAGC AAGGAAGAGA 300

GACGCACATG AACACCATGA TGATGCTATC AGGCCTGATG GAGGGAGCAA CCATGCACCT 360

TTTCCCCTCT GGAAATTCAT AGCTCACACT TTTTTTTAAT GGAAGCAAGA GTTGGCAAAC 420

ACATGCATTT TCAAACAAGG AAAATTAATT CTCAAACCAC CATGACATGC AATTCTCAAA 480

CCATGCACCG ACGAGTCCAT GCGAGGTGGA AACGAAGAAC TGAAAATCAA CATCCCAGTT 540

GTCGAGTCGA GAAGAGGATG ACACTGAAAG TATGCGTATT ACGATTTCAT TTACATACAT 600

GTACAAATAC ATAATGTACC CTACAATTTG TTTTTTGGAG CAGAGTGGTG TGGTCTTTTT 660

TTTTTACACG AAAATGCCAT AGCTGGCCCG CATGCGTGCA GATCGGATGA TCGGTCGGAG 720

ACGACGGACA ATCAGACACT CACCAACTGC TTTTGTCTGG GACACAATAA ATGTTTTTGT 780

AAACAAAATA AATACTTATA AACGAGGGTA CTAGAGGCCG CTAACGGCAT GGCCAGGTAA 840

ACGCGCTCCC AGCCGTTGGT TTGCGATCTC GTCCTCCCGC ACGCAGCGTC GCCTCCACCG 900

TCCGTCCGTC GCTGCCACCT CTGCTGTGCG CGCGCACGAA GGGAGGAAGA ACGAACGCCG 960
CACACACACT CACACACGGC ACACTCCCCG TGGGTCCCCT TTCCGGCTTG GCGTCTATCT 1020
CCTCTCCCCC GCCCATCCCC ATGCACTGCA CCGTACCCGC CAGCTTCCAC CCCCGCCGCA 1080
CACGTTGCTC CCCCTTCTCA TCGCTTCTCA ATTAATATCT CCATCACTCG GGTTCCGCGC 1140
TGCATTTCGG CCGGCGGGTT GAGTGAGATC TGGGCGACTG GCTGACTCAA TCACTACGCG 1200
GGGATGGCGA CGTTCGCGGT GTCCGGCGCG ACTCTCGGTG TGGCGCGGGC CGGCGTCGGA 1260
GTGGCGCGGG CCGGCTCGGA GCGGAGGGGC GGGGCGGACT TGCCGTCGCT GCTCCTCAGG 1320
AAGAAGGACT CCTCTCGTAC GCCTCGCTCT CTCGAATCTC CCCCGTCTGG CTTTGGCTCC 1380
CCTTCTCTCT CCTCTGCGCG CGCATGGCCT GTTCGATGCT GTTCCCCAAT TGATCTCCAT 1440
GAGTGAGAGA GATAGCTGGA TTAGGCGATC GCGCTTCCTG AACCTGTATT TTTTCCCCCG 1500
CGGGGAAATG CGTTAGTGTC ACCCAGGCCC TGGTGTTACC ACGGCTTTGA TCATTCCTCG 1560
TTTCATTCTG ATATATATTT TCTCATTCTT TTTCTTCCTG TTCTTGCTGT AACTGCAAGT 1620
TGTGGCGTTT TTTCACTATT GTAGTCATCC TTGCATTTTG CAGGCGCCGT CCTGAGCCGC 1680
GCGGCCTCTC CAGGGAAGGT CCTGGTGCCT GACGGCGAGA GGACGACTTG GCAAGTCCGG 1740
CGCAACCTGA AGAATTACAG GTACACACAC TCGTGCCGGT AAATCTTCAT ACAATCGTTA 1800
TTCACTTACC AAATGCCGGA TGAAACCAAC CACGGATGCG TCAGGTTTCG AGCTTCTTCT 1860
ATCAGCATTG TGCAGTACTG CACTGCCTTG TTCATTTTGT TAGCCTTGGC CCCGTGCTGG 1920
CTCTTGGGCC ACTGAAAAAA TCAGATGGAT GTGCATTCTA GCAAGAACTT CACAACATAA 1980
TGCACCGTTT GGGGTTTCGT CAGTCTGCTC TACAATTGCT ATTTTTCGTG CTGTAGATAC 2040
CTGAAGATAT CGAGGAGCAA ACGGCGGAAG TGAACATGAC AGGGGGGACT GCAGAGAAAC 2100
TTCAATCTTC AGAACCGACT CAGGGCATTG TGGAAACAAT CACTGATGGT GTAACCAAAG 2160
GAGTTAAGGA ACTAGTCGTG GGGGAGAAAC CGCGAGTTGT CCCAAAACCA GGAGATGGGC 2220
AGAAAATATA CGAGATTGAC CCAACACTGA AAGATTTTCG GAGCCATCTT GACTACCGGT 2280
AATGCCTACC CGCTGCTTTC GCTCATTTTG AATTAAGGTC CTTTCATCAT GCAAATTTGG 2340
GGAACATCAA AGAGACAAAG ACTAGGGACC ACCATTTCAT ACAGATCCCT TCGTGGTCTG 2400
AGAATATGCT GGGAAGTAAA TGTATAATTG ATGGCTACAA TTTGCTCAAA ATTGCAATAC 2460
GAATAACTGT CTCCGATCAT TACAATTAAA GAGTGGCAAA CTGATGAAAA TGTGGTGGAT 2520
GGGTTATAGA TTTTACTTTG CTAATTCCTC TACCAAATTC CTAGGGGGGA AATCTACCAG 2580
TTGGGAAACT TAGTTTCTTA TCTTTGTGGC CTTTTTGTTT TGGGGAAAAC ACATTGCTAA 2640
ATTCGAATGA TTTTGGGTAT ACCTCGGTGG ATTCAACAGA TACAGCGAAT ACAAGAGAAT 2700
TCGTGCTGCT ATTGACCAAC ATGAAGGTGG ATTGGAAGCA TTTTCTCGTG GTTATGAAAA 2760
GCTTGGATTT ACCCGCAGGT AAATTTAAAG CTTTATTATT ATGAAACGCC TCCACTAGTC 2820
TAATTGCATA TCTTATAAGA AAATTTATAA TTCCTGTTTT CCCCTCTCTT TTTTCCAGTG 2880
CTGAAGGTAT CGTCTAATTG CATATCTTAT AAGAAAATTT ATATTCCTGT TTTCCCCTAT 2940

TTTCCAGTGC TGAAGGTATC ACTTACCGAG AATGGGCTCC CTGGAGCGCA TGTTATGTTC 3000
TTTTAAGTTC CTTAACGAGA CACCTTCCAA TTTATTGTTA ATGGTCACTA TTCACCAACT 3060
AGCTTACTGG ACTTACAAAT TAGCTTACTG AATACTGACC AGTTACTATA AATTTATGAT 3120
CTGGCTTTTG CACCCTGTTA CAGTCTGCAG CATTAGTAGG TGACTTCAAC AATTGGAATC 3180
CAAATGCAGA TACTATGACC AGAGTATGTC TACAGCTTGG CAATTTTCCA CCTTTGCTTC 3240
ATAACTACTG ATACATCTAT TTGTATTTAT TTAGCTGTTT GCACATTCCT TAAAGTTGAG 3300
CCTCAACTAC ATCATATCAA AATGGTATAA TTTGTCAGTG TCTTAAGCTT CAGCCCAAAG 3360
ATTCTACTGA ATTTAGTCCA TCTTTTTGAG ATTGAAAATG AGTATATTAA GGATGAATGA 3420
ATACGTGCAA CACTCCCATC TGCATTATGT GTGCTTTTCC ATCTACAATG AGCATATTTC 3480
CATGCTATCA GTGAAGGTTT GCTCCTATTG ATGCAGATAT TTGATATGGT CTTTTCAGGA 3540
TGATTATGGT GTTTGGGAGA TTTTCCTCCC TAACAACGCT GATGGATCCT CAGCTATTCC 3600
TCATGGCTCA CGTGTAAAGG TAAGCTGGCC AATTATTTAG TCGAGGATGT AGCATTTTCG 3660
AACTCTGCCT ACTAAGGGTC CCTTTTCCTC TCTGTTTTTT AGATACGGAT GGATACTCCA 3720
TCCGGTGTGA AGGATTCAAT TTCTGCTTGG ATCAAGTTCT CTGTGCAGGC TCCAGGTGAA 3780
ATACCTTTCA ATGGCATATA TTATGATCCA CCTGAAGAGG TAAGTATCGA TCTACATTAC 3840
ATTATTAAAT GAAATTTCCA GTGTTACAGT TTTTTAATAC CCACTTCTTA CTGACATGTG 3900
AGTCAAGACA ATACTTTTGA ATTTGGAAGT GACATATGCA TTAATTCACC TTCTAAGGGC 3960
TAAGGGGCAA CCAACCTTGG TGATGTGTGT ATGCTTGTGT GTGACATAAG ATCTTATAGC 4020
TCTTTTATGT GTTCTCTGTT GGTTAGGATA TTCCATTTTG GCCTTTTGTG ACCATTTACT 4080
AAGGATATTT ACATGCAAAT GCAGGAGAAG TATGTCTTCC AACATCTCAA CTAAACGACC 4140
AGAGTCACTA AGGATTTATG AATCACACAT TGGAATGAGC AGCCCGGTAT GTCAATAAGT 4200
TATTTCACCT GTTTCTGGTC TGATGGTTTA TTCTATGGAT TTTCTAGTTC TGTTATGTAC 4260
TGTTAACATA TTACATGGTG CATTCACTTG ACAACCTCGA TTTTATTTTC TAATGTCTTC 4320
ATATTGGCAA GTGCAAAACT TTGCTTCCTC TTTGTCTGCT TGTTCTTTTG TCTTCTGTAA 4380
GATTTCCATT GCATTTGGAG GCAGTGGGCA TGTGAAAGTC ATATCTATTT TTTTTTTGTC 4440
AGAGCATAGT TATATGAATT CCATTGTTGT TGCAATAGCT CGGTATAATG TAACCATGTT 4500
ACTAGCTTAA GATTTCCCAC TTAGGATGTA AGAAATATTG CATTGGAGCG TCTCCAGCAA 4560
GCCATTTCCT ACCTTATTAA TGAGAGAGAG ACAAGGGGGG GGGGGGGGGG GGGGTTCCCT 4620
TCATTATTCT GCGAGCGATT CAAAAACTTC CATTGTTCTG AGGTGTACGT ACTGCAGGGA 4680
TCTCCCATTA TGAAGAGGAT ATAGTTAATT CTTTGTAACC TACTTGGAAA CTTGAGTCTT 4740
GAGGCATCGC TAATATATAC TATCATCACA ATACTTAGAG GATGCATCTG AAATTTTAGT 4800
GTGATCTTGC ACAGGAACCG AAGATAAATT CATATGCTAA TTTTAGGGAT GAGGTGTTGC 4860
CAAGAATTAA AAGGCTTGGA TACAATGCAG TGCAGATAAT GGCAATCCAG GAGCATTCAT 4920

ACTATGCAAG CTTTGGGTAT TCACACAATC CATTTTTTTC TGTATACACT CTTCACCCAT 4980
TTGGAGCTAT TACATCCTAA TGCTTCATGC ACATAAAATA TTTGGATATA ATCCTTTATT 5040
AGATATATAG TACAACTACA CTTAGTATTC TGAAAAAGAT CATTTTATTG TTGTTGGCTT 5100
GTTCCAGGTA CCATGTTACT AATTTTTTTG CACCAAGTAG CCGTTTTGGA ACTCCAGAGG 5160
ACTTAAAATC CTTGATCGAT AGAGCACATG AGCTTGGTTT GCTTGTTCTT ATGGATATTG 5220
TTCATAGGTA ATTAGTCCAA TTTAATTTTA GCTGTTTTAC TGTTTATCTG GTATTCTAAA 5280
GGGAAATTCA GGCAATTATG ATACATTGTC AAAAGCTAAG AGTGGCGAAA GTGAAATGTC 5340
AAAATCTAGA GTGGCATAAG GAAAATTGGC AAAAACTAGA GTGGCAAAAA TAAAATTTTC 5400
CCATCCTAAA TGGCAGGGCC CTATCGCCGA ATATTTTTCC ATTCTATATA ATTGTGCTAC 5460
GTGACTTCTT TTTTCTCAGA TGTATTAAAC CAGTTGGACA TGAAATGTAT TTGGTACATG 5520
TAGTAAACTG ACAGTTCCAT AGAATATCGT TTTGTAATGG CAACACAATT TGATGCCATA 5580
GATGTGGATT GAGAAGTTCA GATGCTATCA ATAGAATTAA TCAACTGGCC ATGTACTCGT 5640
GGCACTACAT ATAGTTTGCA AGTTGGAAAA CTGACAGCAA TACCTCACTG ATAAGTGGCC 5700
AGGCCCCACT TGCCAGCTTC ATACTAGATG TTACTTCCCT GTTGAATTCA TTTGAACATA 5760
TTACTTAAAG TTCTTCATTT GTCCTAAGTC AAACTTCTTT AAGTTTGACC AAGTCTATTG 5820
GAAAATATAT CAACATCTAC AACACCAAAT TACTTTGATC AGATTAACAA TTTTTATTTT 5880
ATTATATTAG CACATCTTTG ATGTTGTAGA TATCAGCACA TTTTTCTATA GACTTGGTCA 5940
AATATAGAGA AGTTTGACTT AGGACAAATC TAGAACTTCA ATCAATTTGG ATCAGAGGGA 6000
ACATCAAATA ATATAGATAG ATGTCAACAC TTCAACAAAA AAATCAGACC TTGTCACCAT 6060
ATATGCATCA GACCATCTGT TTGCTTTAGC CACTTGCTTT CATATTTATG TGTTTGTACC 6120
TAATCTACTT TTCCTTCTAC TTGGTTTGGT TGATTCTATT TCAGTTGCAT TGCTTCATCA 6180
ATGATTTTGT GTACCCTGCA GTCATTCGTC AAATAATACC CTTGACGGTT TGAATGGTTT 6240
CGATGGCACT GATACACATT ACTTCCACGG TGGTCCACGC GGCCATCATT GGATGTGGGA 6300
TTCTCGTCTA TTCAACTATG GGAGTTGGGA AGTATGTAGC TCTGACTTCT GTCACCATAT 6360
TTGGCTAACT GTTCCTGTTA ATCTGTTCTT ACACATGTTG ATATTCTATT CTTATGCAGG 6420
TATTGAGATT CTTACTGTCA AACGCGAGAT GGTGGCTTGA AGAATATAAG TTTGATGGAT 6480
TTCGATTTGA TGGGGTGACC TCCATGATGT ATACTCACCA TGGATTACAA GTAAGTCATC 6540
AAGTGGTTTC AGTAACTTTT TTAGGGCACT GAAACAATTG CTATGCATCA TAACATGTAT 6600
CATGATCAGG ACTTGTGCTA CGGAGTCTTA GATAGTTCCC TAGTATGCTT GTACAATTTT 6660
ACCTGATGAG ATCATGGAAG ATTGGAAGTG ATTATTATTT ATTTTCTTTC TAAGTTTGTT 6720
TCTTGTTCTA GATGACATTT ACTGGGAACT ATGGCGAATA TTTTGGATTT GCTACTGATG 6780
TTGATGCGGT AGTTTACTTG ATGCTGGTCA ACGATCTAAT TCATGGACTT TATCCTGATG 6840
CTGTATCCAT TGGTGAAGAT GTAAGTGCTT ACAGTATTTA TGATTTTTAA CTAGTTAAGT 6900
AGTTTTATTT TGGGGATCAG TCTGTTACAC TTTTTGTTAG GGGTAAAATC TCTCTTTTCA 6960

TAACAATGCT AATTTATACC TTGTATGATA ATGCATCACT TAGTAATTTG AAAAGTGCAA 7020
GGGCATTCAA GCTTACGAGC ATATTTTTTG ATGGCTGTAA TTTATTTGAT AGTATGCTTG 7080
TTTGGGTTTT TCAATAAGTG GGAGTGTGTG ACTAATGTTG TATTATTTAT TTAATTGCGG 7140
AAGAAATGGG CAACCTTGTC AATTGCTTCA GAAGGCTAAC TTTGATTCCA TAAACGCTTT 7200
GGAAATGAGA GGCTATTCCC AAGGACATGA ATTATACTTC AGTGTGTTCT GTACATGTAT 7260
TTGTAATAGT GGTTTAACTT AAATTCCTGC ACTGCTATGG AATCTCACTG TATGTTGTAG 7320
TGTACACATC CACAAACAAG TAATCCTGAG CTTTCAACTC ATGAGAAAAT AGAGTCCGCT 7380
TCTGCCAGCA TTAACTGTTC ACAGTTCTAA TTTGTGTAAC TGTGAAATTG TTCAGGTCAG 7440
TGGAATGCCT ACATTTTGCA TCCCTGTTCC AGATGGTGGT GTTGGTTTTG ACTACCGCCT 7500
GCATATGGCT GTAGCAGATA AATGGATTGA ACTCCTCAAG TAAGTGCAGG AATATTGGTG 7560
ATTACATGCG CACAATGATC TAGATTACAT TTTCTAAATG GTAAAAAGGA AAATATGTAT 7620
GTGAATATCT AGACATTTGC CTGTTATCAG CTTGAATACG AGAAGTCAAA TACATGATTT 7680
AAATAGCAAA TCTCGGAAAT GTAATGGCTA GTGTCTTTAT GCTGGGCAGT GTACATTGCG 7740
CTGTAGCAGG CCAGTCAACA CAGTTAGCAA TATTTTCAGA AACAATATTA TTTATATCCG 7800
TATATGAGAA AGTTAGTATA TAAACTGTGG TCATTAATTG TGTTCACCTT TTGTCCTGTT 7860
TAAGGATGGG CAGTAGGTAA TAAATTTAGC CAGATAAAAT AAATCGTTAT TAGGTTTACA 7920
AAAGGAATAT ACAGGGTCAT GTAGCATATC TAGTTGTAAT TAATGAAAAG GCTGACAAAA 7980
GGCTCGGTAA AAAAAACTTT ATGATGATCC AGATAGATAT GCAGGAACGC GACTAAAGCT 8040
CAAATACTTA TTGCTACTAC ACAGCTGCCA ATCTGTCATG ATCTGTGTTC TGCTTTGTGC 8100
TATTTAGATT TAAATACTAA CTCGATACAT TGGCAATAAT AAACTTAACT ATTCAACCAA 8160
TTTGGTGGAT ACCAGAATTT CTGCCCTCTT GTTAGTAATG ATGTGCTCCC TGCTGCTGTT 8220
CTCTGCCGTT ACAAAAGCTG TTTTCAGTTT TTTGCATCAT TATTTTTGTG TGTGAGTAGT 8280
TTAAGCATGT TTTTTGAAGC TGTGAGCTGT TGGTACTTAA TACATTCTTG GAAGTGTCCA 8340
AATATGCTGC AGTGTAATTT AGCATTTCTT TAACACAGGC AAAGTGACGA ATCTTGGAAA 8400
ATGGGCGATA TTGTGCACAC CCTAACAAAT AGAAGGTGGC TTGAGAAGTG TGTAACTTAT 8460
GCAGAAAGTC ATGATCAAGC ACTAGTTGGT GACAAGACTA TTGCATTCTG GTTGATGGAT 8520
AAGGTACTAG CTGTTACTTT TGGACAAAAG AATTACTCCC TCCCGTTCCT AAATATAAGT 8580
CTTTGTAGAG ATTCCACTAT GGACCACATA GTATATAGAT GCATTTTAGA GTGTAGATTC 8640
ACTCATTTTG CTTCGTATGT AGTCCATAGT GAAATCTCTA CAGAGACTTA TATTTAGGAA 8700
CGGAGGGAGT ACATAATTGA TTTGTCTCAT CAGATTGCTA GTGTTTTCTT GTGATAAAGA 8760
TTGGCTGCCT CACCCATCAC CAGCTATTTC CCAACTGTTA CTTGAGCAGA ATTTGCTGAA 8820
AACGTACCAT GTGGTACTGT GGCGGCTTGT GAACTTTGAC AGTTATGTTG CAATTTTCTG 8880
TTCTTATTTA TTTGATTGCT TATGTTACCG TTCATTTGCT CATTCCTTTC CGAGACCAGC 8940

CAAAGTCACG TGTTAGCTGT GTGATCTGTT ATCTGAATCT TGAGCAAATT TTATTAATAG 9000
GCTAAAATCC AACGAATTAT TTGCTTGAAT TTAAATATAC AGACGTATAG TCACCTGGCT 9060
CTTTCTTAGA TGATTACCAT AGTGCCTGAA GGCTGAAATA GTTTTGGTGT TTCTTGGATG 9120
CCGCCTAAAG GAGTGATTTT TATTGGATAG ATTCCTGGCC GAGTCTTCGT TACAACATAA 9180
CATTTTGGAG ATATGCTTAG TAACAGCTCT GGGAAGTTTG GTCACAAGTC TGCATCTACA 9240
CGCTCCTTGA GGTTTTATTA TGGCGCCATC TTTGTAACTA GTGGCACCTG TAAGGAAACA 9300
CATTCAAAAG GAAACGGTCA CATCATTCTA ATCAGGACCA CCATACTAAG AGCAAGATTC 9360
TGTTCCAATT TTATGAGTTT TTGGGACTCC AAAGGGAACA AAAGTGTCTC ATATTGTGCT 9420
TATAACTACA GTTGTTTTTA TACCAGTGTA GTTTTATTCC AGGACAGTTG ATACTTGGTA 9480
CTGTGCTGTA AATTATTTAT CCGACATAGA ACAGCATGAA CATATCAAGC TCTCTTTGTG 9540
CAGGATATGT ATGATTTCAT GGCTCTGGAT AGGCTTCAAC TCTTCGCATT GATCGTGGCA 9600
TAGCATTACA TAAAATGATC AGGCTTGTCA CCATGGGTTT AGGTGGTGAA GGCTATCTTA 9660
ACTTCATGGG AAATGAGTTT GGGCATCCTG GTCAGTCTTT ACAACATTAT TGCATTCTGC 9720
ATGATTGTGA TTTACTGTAA TTTGAACCAT GCTTTTCTTT CACATTGTAT GTATTATGTA 9780
ATCTGTTGCT TCCAAGGAGG AAGTTAACTT CTATTTACTT GGCAGAATGG ATAGATTTTC 9840
CAAGAGGCCC ACAAACTCTT CCAACCGGCA AAGTTCTCCC CTGGAAATAA CAATAGTTAT 9900
GATAAATGCC GCCGTAGATT TGATCTTGTA AGTTTTAGCT GTGCTATTAC ATTCCCTCAC 9960
TAGATCTTTA TTGGCCATTT ATTTCTTGAT GAAATCATAA TGTTTGTTAG GAAAGATCAA 10020
CATTGCTTTT GTAGTTTTGT AGACGTTAAC ATAAGTATGT GTTGAGAGTT GTTGATCATT 10080
AAAAATATCA TGATTTTTTG CAGGGAGATG CAGATTTTCT TAGATATCGT GGTATGCAAG 10140
AGTTCGATCA GGCAATGCAG CATCTTGAGG AAAAATATGG GGTATGTCAC TGGTTTGTCT 10200
TTGTTGCATA ACAAGTCACA GTTTAACGTC AGTCTCTTCA AGTGGTAAAA AAAGTGTAGA 10260
ATTAATTCCT GTAATGAGAT GAAAACTGTG CAAAGGCGGA GCTGGAATTG CTTTTCACCA 10320
AAACTATTTT CTTAAGTGCT TGTGTATTGA TACATATACC AGCACTGACA ATGTAACTGC 10380
AGTTTATGAC ATCTGAGCAC CAGTATGTTT CACGGAAACA TGAGGAAGAT AAGGTGATCA 10440
TCCTCAAAAG AGGAGATTTG GTATTTGTTT TCAACTTCCA CTGGAGCAAT AGCTTTTTTG 10500
ACTACCGTGT TGGGTGTTCC AAGCCTGGGA AGTACAAGGT ATGCTTGCCT TTTCATTGTC 10560
CACCCTTCAC CAGTAGGGTT AGTGGGGGCT TCTACAACTT TTAATTCCAC ATGGATAGAG 10620
TTTGTTGGTC GTGCAGCTAT CAATATAAAG AATAGGGTAA TTTGTAAAGA AAAGAATTTG 10680
CTCGAGCTGT TGTAGCCATA GGAAGGTTGT TCTTAACAGC CCCGAAGCAC ATACCATTCA 10740
TTCATATTAT CTACTTAAGT GTTTGTTTCA ATCTTTATGC TCAGTTGGAC TCGGTCTAAT 10800
ACTAGAACTA TTTTCCGAAT CTACCCTAAC CATCCTAGCA GTTTTAGAGC AGCCCCATTT 10860
GGACAATTGG CTGGGTTTTT GTTAGTTGTG ACAGTTTCTG CTATTTCTTA ATCAGGTGGC 10920
CTTGGACTCT GACGATGCAC TCTTTGGTGG ATTCAGCAGG CTTGATCATG ATGTCGACTA 10980

CTTCACAACC GTAAGTCTGG GCTCAAGCGT CACTTGACTC GTCTTGACTC AACTGCTTAC 11040

AAATCTGAAT CAACTTCCCA ATTGCTGATG CCCTTGCAGG AACATCCGCA TGACAACAGG 11100

CCGCGCTCTT TCTCGGTGTA CACTCCGAGC AGAACTGCGG TCGTGTATGC CCTTACAGAG 11160

TAAGAACCAG CAGCGGCTTG TTACAAGGCA AAGAGAGAAC TCCAGAGAGC TCGTGGATCG 11220

TGAGCGAAGC GACGGGCAAC GGCGCGAGGC TGCTCCAAGC GCCATGACTG GGAGGGGATC 11280

GTGCCTCTTC CCCAGATGCC AGGAGGAGCA GATGGATAGG TAGCTTGTTG GTGAGCGCTC 11340

GAAAGAAAAT GGACGGGCCT GGGTGTTTGT TGTGCTGCAC TGAACCCTCC TCCTATCTTG 11400

CACATTCCCG GTTGTTTTTG TACATATAAC TAATAATTGC CCGTGCGCTC AACGTGAAAA 11460

TCC 11463

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2662 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE:

(vi) ORIGINAL SOURCE:

(A) ORGANISM: triticum tauschii

(F) TISSUE TYPE: Endosperm

(ix) FEATURE:

(A) NAME/KEY: misc_feature

(B) LOCATION:1..2651

(D) OTHER INFORMATION:/product= "nucleotide sequence of cDNA wheat SSS I"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

TCTCCCACTC TTCTCTCCCC GCGCACACCG AGTCGGCACC GGCTCATCAC CCATCACCTC 60

GGCCTCGGCC ACCGGCAAAC CCCCCGATCC GCTTTTGCAG GCAGCGCACT AAAACCCCGG 120

GGAGCGCGCC CCGCGGCAGC AGCAGCACCG CAGTGGGAGA GAGAGGCTTC GCCCCGGCCC 180

GCACCGAGCG GGGCGATCCA CCGTCCGTGC GTCCGCACCT CCTCCGCCTC CTCCCCTGTC 240

CCGCGCGCCC ACACCCATGG CGGCGACGGG CGTCGGCGCC GGGTGCCTCG CCCCCAGCGT 300

CCGCCTGCGC GCCGATCCGG CGACGGCGGC CCGGGCGTCC GCCTGCGTCG TCCGCGCGCG 360

GCTCCGGCGC TTGGCGCGGG GCCGCTACGT TGCCGAGCTC AGCAGGGAGG GCCCCGCGGC 420

GCGCCCCGCG CAGCAGCAGC AACTGGCCCC GCCGCTCGTG CCAGGCTTCC TCGCGCCGCC 480

GCCGCCCGCG CCCGCCCAGT CGCCGGCCCC GACGCAGCCG CCCCTGCCGG ACGCCGGCGT 540

GGGGGAACTC GCGCCCGACC TCCTGCTCGA AGGGATTGCT GAGGATTCCA TCGACAGCAT 600

AATTGTGGCT GCAAGTGAGC AGGATTCTGA GATCATGGAT GCGAATGAGC AACCTCAAGC 660
TAAAGTTACA CGTAGCATCG TGTTTGTGAC TGGTGAAGCT GCTCCTTATG CAAAGTCAGG 720
GGGGCTGGGA GATGTTTGTG GTTCGTTACC AATTGCTCTT GCTGCTCGTG GTCACCGTGT 780
GATGGTTGTA ATGCCAAGAT ACTTGAATGG GTCCTCTGAT AAAAACTATG CAAAGGCATT 840
ATACACTGGG AAGCACATTA AGATTCCATG CTTTGGGGGA TCACATGAAG TGACCTTTTT 900
TCATGAGTAT AGAGACAACG TCGATTGGGT GTTTGTCGAT CATCCGTCAT ATCATAGACC 960
AGGAAGTTTA TATGGAGATA ATTTTGGTGC TTTTGGTGAT AATCAGTTCA GATACACACT 1020
CCTTTGCTAT GCTGCATGCG AGGCCCCACT AATCCTTGAA TTGGGAGGAT ATATTTATGG 1080
ACAGAATTGC ATGTTTGTTG TGAACGATTG GCATGCCAGC CTTGTGCCAG TCCTTCTTGC 1140
TGCAAAATAT AGACCATACG GTGTTTACAG AGATTCCCGC AGCACCCTTG TTATACATAA 1200
TTTAGCACAT CAGGGTCTGG AGCCTGCAAG TACATATCCT GATCTGGGAT TGCCACCTGA 1260
ATGGTATGGA GCTTTAGAAT GGGTATTTCC AGAATGGGCA AGGAGGCATG CCCTTGACAA 1320
GGGTGAGGCA GTTAACTTTT TGAAAGGAGC AGTCGTGACA GCAGATCGAA TTGTGACCGT 1380
CAGTCAGGGT TATTCATGGG AGGTCACAAC TGCTGAAGGT GGACAGGGCC TCAATGAGCT 1440
CTTAAGCTCC CGAAAAAGTG TATTGAATGG AATTGTAAAT GGAATTGACA TTAATGATTG 1500
GAACCCCACC ACAGACAAGT GTCTCCCTCA TCATTATTCT GTCGATGACC TCTCTGGAAA 1560
GGCCAAATGT AAAGCTGAAT TGCAGAAGGA GCTGGGTTTA CCTGTAAGGG AGGATGTTCC 1620
TCTGATTGGC TTTATTGGAA GACTGGATTA CCAGAAAGGC ATTGATCTCA TTAAAATGGC 1680
CATTCCAGAG CTCATGAGGG AGGACGTGCA GTTTGTCATG CTTGGATCTG GGGATCCAAT 1740
TTTTGAAGGC TGGATGAGAT CTACCGAGTC GAGTTACAAG GATAAATTCC GTGGATGGGT 1800
TGGATTTAGT GTTCCAGTTT CCCACAGAAT AACTGCAGGT TGCGATATAT TGTTAATGCC 1860
ATCCAGGTTT GAACCTTGTG GTCTTAATCA GCTATATGCT ATGCAATATG GTACAGTTCC 1920
TGTAGTTCAT GGAACTGGGG GCCTCCGAGA CACAGTCGAG ACCTTCAACC CTTTTGGTGC 1980
AAAAGGAGAG GAGGGTACAG GGTGGGCGTT CTCACCGCTA ACCGTGGACA AGATGTTGTG 2040
GGCATTGCGA ACCGCGATGT CGACATTCAG GGAGCACAAG CCGTCCTGGG AGGGGCTCAT 2100
GAAGCGAGGC ATGACGAAAG ACCATACGTG GGACCATGCC GCCGAGCAGT ACGAGCAGAT 2160
CTTCGAATGG GCCTTCGTGG ACCAACCCTA CGTCATGTAG ACGGGGACTG GGGAGGTCGA 2220
AGCGCGGGTC TCCTTGAGCT CTGAAGACAT GTTCCTCATC CTTCCGCGGC CCGGAAGGAT 2280
ACCCCTGTAC ATTGCGTTGT CCTGCTACAG TAGAGTCGCA ATGCGCCTGC TTGCTTGGTC 2340
CGCCGGTTCG AGAGTAGATG ACGGCTGTGC TGCTGCGGCG GTGACAGCTT CGGGTGGATG 2400
ACAGTTACAG TTTTGGGGAA TAAGGAAGGG ATGTGCTGCA GGATGGTTAA CAGCAAAGCA 2460
CCACTCAGAT GGCAGCCTCT CTGTCCGTGT TACAGCTGAA ATCAGAAACC AACTGGTGAC 2520
TCTTTAGCCT TAGCGATTGT GAAGTTTGTT GCATTCTGTG TATGTTGTCT TGTCCTTAGC 2580

TGACAAATAT TAGACCTGTT GGAGAATTTT ATTTATCTTT GCTGCTGTTG TTTTTGTTTT 2640

GTTAAAAAAA AAAAAAAAAA AA 2662

(2) INFORMATION FOR SEQ ID NO: 12:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 768 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii

(ix) FEATURE:
(A) NAME/KEY: Protein
(B) LOCATION:1..768

(ix) FEATURE:
(A) NAME/KEY: Protein
(B) LOCATION:1..768
(D) OTHER INFORMATION:/product= "deduced amino acid sequence SBE II"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

```
Met Ala Thr Phe Ala Val Ser Gly Ala Thr Leu Gly Val Ala Arg Pro
1               5                   10                  15

Pro Ala Ala Ala Gln Pro Glu Glu Leu Gln Ile Pro Glu Asp Ile Glu
            20                  25                  30

Glu Gln Thr Ala Glu Val Asn Met Thr Gly Gly Thr Ala Glu Lys Leu
        35                  40                  45

Glu Ser Ser Glu Pro Thr Gln Gly Ile Val Glu Thr Ile Thr Asp Gly
    50                  55                  60

Val Thr Lys Gly Val Lys Glu Leu Val Val Gly Glu Lys Pro Arg Val
65                  70                  75                  80

Val Pro Lys Pro Gly Asp Gly Gln Lys Ile Tyr Glu Ile Asp Pro Thr
                85                  90                  95

Leu Lys Asp Phe Arg Ser His Leu Asp Tyr Arg Tyr Ser Glu Tyr Arg
            100                 105                 110

Arg Ile Arg Ala Ala Ile Asp Gln His Glu Gly Gly Leu Glu Ala Phe
        115                 120                 125

Ser Arg Gly Tyr Glu Lys Leu Gly Phe Thr Arg Ser Ala Glu Gly Ile
    130                 135                 140

Thr Tyr Arg Glu Trp Ala Pro Gly Ala His Ser Ala Ala Leu Val Gly
145                 150                 155                 160
```

```
Asp Phe Asn Asn Trp Asn Pro Asn Ala Asp Thr Met Thr Arg Asp Asp
                165                 170                 175

Tyr Gly Val Trp Glu Ile Phe Leu Pro Asn Asn Ala Asp Gly Ser Pro
            180                 185                 190

Ala Ile Pro His Gly Ser Arg Val Lys Ile Arg Met Asp Thr Pro Ser
        195                 200                 205

Gly Val Lys Asp Ser Ile Ser Ala Trp Ile Lys Phe Ser Val Gln Ala
    210                 215                 220

Pro Gly Glu Ile Pro Phe Asn Gly Ile Tyr Tyr Asp Pro Pro Glu Glu
225                 230                 235                 240

Glu Lys Tyr Val Phe Gln His Pro Gln Pro Lys Arg Pro Glu Ser Leu
                245                 250                 255

Arg Ile Tyr Glu Ser His Ile Gly Met Ser Ser Pro Glu Pro Lys Ile
            260                 265                 270

Asn Ser Tyr Ala Asn Phe Arg Asp Glu Val Leu Pro Arg Ile Lys Arg
        275                 280                 285

Leu Gly Tyr Asn Ala Val Gln Ile Met Ala Ile Gln Glu His Ser Tyr
    290                 295                 300

Tyr Ala Ser Phe Gly Tyr His Val Thr Asn Phe Phe Ala Pro Ser Ser
305                 310                 315                 320

Arg Phe Gly Thr Pro Glu Asp Leu Lys Ser Leu Ile Asp Arg Ala His
                325                 330                 335

Glu Leu Gly Leu Leu Val Leu Met Asp Ile Val His Ser His Ser Ser
            340                 345                 350

Asn Asn Thr Leu Asp Gly Leu Asn Gly Phe Asp Gly Thr Asp Thr His
        355                 360                 365

Tyr Phe His Gly Gly Pro Arg Gly His His Trp Met Trp Asp Ser Arg
    370                 375                 380

Leu Phe Asn Tyr Gly Ser Trp Glu Val Leu Arg Phe Leu Leu Ser Asn
385                 390                 395                 400

Ala Arg Trp Trp Leu Glu Glu Tyr Lys Phe Asp Gly Phe Arg Phe Asp
                405                 410                 415

Gly Val Thr Ser Met Met Tyr Thr His His Gly Leu Gln Met Thr Phe
            420                 425                 430

Thr Gly Asn Tyr Gly Glu Tyr Phe Gly Phe Ala Thr Asp Val Asp Ala
        435                 440                 445

Val Val Tyr Leu Met Leu Val Asn Asp Leu Ile His Gly Leu His Pro
    450                 455                 460

Asp Ala Val Ser Ile Gly Glu Asp Val Ser Gly Met Pro Thr Phe Cys
465                 470                 475                 480

Ile Pro Val Pro Asp Gly Gly Val Gly Phe Asp Tyr Arg Leu His Met
                485                 490                 495
```

```
Ala Val Ala Asp Lys Trp Ile Glu Leu Leu Lys Gln Ser Asp Glu Ser
            500                 505                 510

Trp Lys Met Gly Asp Ile Val His Thr Leu Thr Asn Arg Arg Trp Leu
        515                 520                 525

Glu Lys Cys Val Thr Tyr Ala Glu Ser His Asp Gln Ala Leu Val Gly
    530                 535                 540

Asp Lys Thr Ile Ala Phe Trp Leu Met Asp Lys Asp Met Tyr Asp Phe
545                 550                 555                 560

Met Ala Leu Asp Arg Pro Ser Thr Pro Arg Ile Asp Arg Gly Ile Ala
                565                 570                 575

Leu His Lys Met Ile Arg Leu Val Thr Met Gly Leu Gly Gly Glu Gly
            580                 585                 590

Tyr Leu Asn Phe Met Gly Asn Glu Phe Gly His Pro Glu Trp Ile Asp
        595                 600                 605

Phe Pro Arg Gly Pro Gln Thr Leu Pro Thr Gly Lys Val Leu Pro Gly
    610                 615                 620

Asn Asn Asn Ser Tyr Asp Lys Cys Arg Arg Arg Phe Asp Leu Gly Asp
625                 630                 635                 640

Ala Asp Phe Leu Arg Tyr His Gly Met Gln Glu Phe Asp Gln Ala Met
                645                 650                 655

Gln His Leu Glu Glu Lys Tyr Gly Phe Met Thr Ser Glu His Gln Tyr
            660                 665                 670

Val Ser Arg Lys His Glu Glu Asp Lys Val Ile Ile Phe Glu Arg Gly
        675                 680                 685

Asp Leu Val Phe Val Phe Asn Phe His Trp Ser Asn Ser Phe Phe Asp
    690                 695                 700

Tyr Arg Val Gly Cys Ser Arg Pro Gly Lys Tyr Lys Val Ala Leu Asp
705                 710                 715                 720

Ser Asp Asp Ala Leu Phe Gly Gly Phe Ser Arg Leu Asp His Asp Val
                725                 730                 735

Asp Tyr Phe Thr Thr Glu His Pro His Asp Asn Arg Pro Arg Ser Phe
            740                 745                 750

Ser Val Tyr Thr Pro Ser Arg Thr Ala Val Val Tyr Ala Leu Thr Glu
        755                 760                 765
```

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10550 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: both

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:1..316
(D) OTHER INFORMATION:/product= "exon 1"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:1472..1828
(D) OTHER INFORMATION:/product= "exon 2"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:2766..2823
(D) OTHER INFORMATION:/product= "exon 3"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:2906..3028
(D) OTHER INFORMATION:/product= "exon 4"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:4113..4194
(D) OTHER INFORMATION:/product= "exon 5"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:4286..4459
(D) OTHER INFORMATION:/product= "exon 6"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:4562..4643
(D) OTHER INFORMATION:/product= "exon 7"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:4744..4855
(D) OTHER INFORMATION:/product= "exon 8"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:4999..5021
(D) OTHER INFORMATION:/product= "exon 9"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:5102..5192
(D) OTHER INFORMATION:/product= "exon 10"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:8593..8718

(D) OTHER INFORMATION:/product= "exon 11"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:8807..8915
(D) OTHER INFORMATION:/product= "exon 12"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:8992..9104
(D) OTHER INFORMATION:/product= "exon 13"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:9161..9199
(D) OTHER INFORMATION:/product= "exon 14"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:9498..9713
(D) OTHER INFORMATION:/product= "exon 15"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATGGCGGCGA CGGGCGTCGG CGCCGGGTGC CTCGCCCCCA GCGTCCGCCT 50

GCGCGCCGAT CCGGCGACGG CGGCCCGGGC GTCCGCTTGC GTCGTCCGCG 100

CGCGGCTCCG GCGCTTGGCG CGGGGCCGCT ACGTCGCCGA GCTCAGCAGG 150

GAGGGCCCCG CGGCGCGCCC CGCGCAGCAG CAGCAACTGG CCCCGCCGCT 200

CGTGCCAGGC TTCCTCGCGC CGCCGCCGCC CGCGCCCGCC CAGTCGCCGG 250

CCCCGACGCA GCCGCCCCTG CCGGACGCCG GCGTGGGGGA ACTCGCGCCC 300

GACCTCCTGC TCGAAGGTAA AAAACAAGGC TGAATCCTCA GATCACTCCG 350

CGTCTTCGTT TTACCAAATA CGGTACTGCG AAGTGGTGCT GTATATGTGA 400

AGTTTCTGTC GATTTCTTCC TGACGGATGT TCAGTCGATT CAGTTGTATA 450

TATGTGATAC GTTCGTTGTT CATCGATCGT ACAGATTTAC CAGCACACTA 500

GATAGAAATC GAGACCGACG CGGGCAGATC AATAGATTTT TCTAGACGTT 550

TTATTGGATC GTGAGATGAT TGATTGGGGT GGCGTGTCGA TACGATAGCG 600

GTGCACCGCC GATGTATCGG GGCATGTGCA CGTGGTTGGG TCTCAGCAGA 650

CATATCACTA GACTGGTATC GTAATTTACT AGTACTACTG GAAAGAGGAC 700

TAAAAAGGCT AGGCCAAGTG CACGCATGTT GGGAACGTTG TTAAATTGAT 750

GAGTTTGTCC TTTGCTTGGG CTGGTATTAT TACCAAAAAA TGGTGTTAGT 800

CCCTGTACTT ATTAATGGGA AAATCTTAAC ATGACACTGG GGTTTATGAG 850
TCTCCAATTG TATATTCTCA GCACTCAACT GATTTTACTG ATACTGTAGT 900
GGAAATGACA CGTGAGCACC CCCCTTCAAG GAATGCAATG CTTCTTTCTG 950
TTTTATATTA CAGGAACTAG AAGGAGCTTC CACCTTTGAG TACAGAAGTA 1000
CTCCCTCCGT TCCAAAATAG ATGACTCAAC TTTGTACTAA TTTTGTACTA 1050
TAGTTAGTAC AAAGTTGAGT CATCTATTTT AGAACGGAGG GAGTAGTATC 1100
GAAATTGAAG ACCCTTGTAT TACTGTCTTG TTTTTCAATG AAAATGGGAG 1150
GCCCATGCAG TAAGTCACAT GGGCACCTGG GAGGCTGGGA TCATGTGTGC 1200
TTTGCAGAGT ACTAGACCCA GCTCACCCTC TGTTAGATTA CTTGTTGGGC 1250
TGCTACTTTG TGTTTGCTGT GCAGTATATC AGACATCCTG AATTTGGCAT 1300
CTAGCTGAGA ACAGAATGCA GGTTGCACCA TTCTTATTAT TGCTAAACTG 1350
TTGTCACGCA ATTTATAAAG AATGTGATCT TCTGAGTATT AATTAATCAT 1400
GTTCTGCTAA TATCTGTCCT CGCTCTGGTG TTGACAAATA TACCATATGA 1450
ATATTTTCCA TTTTGCAACC AGGGATTGCT GAGGATTCCA TCGACAGCAT 1500
AATCGTGGCT GCAAGTGAGC AGGATTCTGA GATCATGGAT GCGAATGAGC 1550
AACCTCAAGC TAAAGTTACA CGTAGCATCG TGTTTGTGAC TGGTGAAGCT 1600
GCTCCTTATG CAAAGTCAGG GGGGCTGGGA GATGTTTGTG GTTCGTTACC 1650
AATTGCTCTT GCTGCTCGTG GTCACCGTGT GATGGTTGTA ATGCCAAGAT 1700
ACTTGAATGG GTCCTCTGAT AAAAACTATG CAAAGGCATT ATACACTGCG 1750
AAGCACATTA AGATTCCATG CTTTGGGGGA TCACATGAAG TGACCTTTTT 1800
TCATGAGTAT AGAGACAACG TCGATTGGGT GGGTACACAA TCACCTTCTT 1850
ATTCTCTGTT GAATTGTAGC AACTGTTTAT CCTTGTTTAC ACTTCTTTTA 1900
GCCCTGCAAA GACATATGTG ATTTCCATAC TTTTTTGTTA TTTCCCTTGT 1950
ACTCTTGCTC ATGAAGGTCA AAATATCATA TATCCATGGA AGTCATGCAT 2000
GTGCCTAGTA TTTTTGGTGT CGGTGCCTTT AACTTTCAGG GATTAATACG 2050
TGGAATTTGA TAACTAAAGT TTATTTTATT GAAAAAAATT GTAGGTTGG 2100
TGAGCCCACA GCCACGCAGT GGCACCACTG CTTGCACATG ATTTTGCATT 2150
TCTGTTTGCA CCGAGCACTT CATGTGAATA AGGTGTAAAA TCATAAAGTA 2200

CCAATTTTAT TCTGCCAATT GCACTTAAGA GTATATACAT TTATCTTGGC 2250
CTCAATCATG GGAGTACTGT GCATTCAGTG CACCATCATT GTTCTAAGGA 2300
GAAAATGTGG GTGCAAGGAA GACACTTTTG TCCCTTAATA AAAGGCAGGC 2350
ACTCTGTTGT CATATAGATA GAAAGCAACA AACTTATTTC AAAGAGCTAA 2400
CAATGGCAAA AGAACCAAAA AAAGCATGCT AAGGCGGTGA CACCAAAAGG 2450
TGAGGGGGGC CTTGTGACTG ACAGCACCCC AAACTATTGC CATTGTTTTA 2500
CTAAATGAAG ATCATTTTAG AAGCTCTCAG GAACTTCGAA AACAGTGGCT 2550
TTCCGTCCAC AGATCGTCTG TTAATATTTT TGTCCAGTGA TACTTTTTTT 2600
GCTCCTTACA AGAGTGCCTA TGTTGACATA TACATTGTTA AGTTGTTCAT 2650
AAGTTTACTT CTTATTCTAA ACAGCAAGTG CCTAATGCTT GCATTTATTT 2700
TGGCTATTTA TTTTTATTCT CATTTCAATC AACACTTTTG TTCAGGTGTT 2750
TGTCGATCAT CCGTCATATC ATAGACCAGG AAGTTTATAT GGAGATAATT 2800
TTGGTGCTTT TGGTGATAAT CAGGTACACT ACACTATACT AAGCTCCTAG 2850
TTGACTAAGT CGTAAGTTGT ACCTCCTCGC TGACCGGCTG CTCTATGTCG 2900
TGCAGTTCAG ATACACACTC CTTTGCTATG CTGCATGCGA GGCCCCACTA 2950
ATCCTTGAAT TGGGAGGATA TATTTATGGA CAGAATTGCA TGTTTGTTGT 3000
GAACGATTGG CATGCCAGCC TTGTGCCAGT GTACGTTGTT TGTGGATCTG 3050
AAAGTCCAAT CCTTTATTCA TTCTCTGCTT TGCAGTGTGC CCATGTCTAC 3100
ATTTCTTTTA TGCTTTTTTC ATGTCTGTTC TTATATTGCA TATATGCTTA 3150
TGGAGTCTAA AAGTTACCGG AGGGAATAAC TCTTAAGGAT TTCCTCAATC 3200
AATTATCTTT AGCTTTAGTT AACATTTACT GTGGCAAACA TAATGTGTTT 3250
TGAGATTTAC AAGTTCAGAG ATTGCACTTC ACTAGTTCGT AGCTAATCTG 3300
ATGTTTTCCC CGAGAAAATG CCTAAAGCTT TGTGTCTTGA TGCATTGATA 3350
GAAAAAGAGT TTATGTACAC TCCCAAAGAG GGGACCCAAA ATTACAACAC 3400
CACACCCCTG AGAACTAGGC GCTGCCGGAA GAAGCGATGC AAGCCCCACT 3450
GCCCCTGCCT TAGCTCAAAG CCGGGCGTCA GCTTGATTGT GTCAAGTAAG 3500
CTAGCAGTGC TAGATTGCGC AAGGTCGATT CGTCGAAGAT GACAGTGTTG 3550
CGCTGCTTCC AAATCCACCA AACTATGAGC ATGATCACTG GAGAAGTACC 3600

TTTTCTCGCG GCTGAGGGGG TGGACTGGTG GTCTGCTGCT GCCAGTTTTC 3650

AGATAATCTG AAAAATGCAT GTTTTGATGA TTTTAGTATC TTGCGGACCC 3700

TGGGTACCAC CTAAGCTTTC ACACAGTAAT TTGCAGTTAC ACCTATAAAA 3750

GTAACGGTCA TGATATGCAT GTGTTTTGGG TAGATCATGG TGCATGCATT 3800

TTAGGAATTA GGACATGCCA GAACCACGTG AGGCTTATGG GGCAATTCAT 3850

TTGTTCCATT ATACGAGTCA TGAATATGGT TCAGCATGTT TGGACGCTAC 3900

TTGTTTGGGG CAATTTCAGA TGGTGAATTG TAGCTGCTTG ATGTTGGCTA 3950

GCTGGCTTAT TTTGTACAAG TATCGATGTT AGATGCATAT TTCCTTTTGT 4000

TCTTGTGCTG TTTGCCATGT TGTATTCCCC TTTTCTGTCG CCAGTGTTGC 4050

ATGTTAAATT GGTTTTCATT ACATAATCAA CTTTGTTGCT GACATCAGTC 4100

ATTTTTATTC AGCCTTCTTG CTGCAAAATA TAGACCATAC GGTGTTTACA 4150

GAGATTCCCG CAGCACCCTT GTTATACATA ATTTAGCACA TCAGGTTTGG 4200

GTCTATCACC TTTCATTATC CGTACATGGC TTTGTAAGTC GGTTCACACG 4250

TATCGTCATA CTGTATGTTA TTTCAATGTC ATTAGGGTGT GGAGCCTGCA 4300

AGTACATATC CTGATCTGGG ATTGCCACCT GAATGGTATG GAGCTTTAGA 4350

ATGGGTATTT CCAGAATGGG CAAGGAGGCA TGCCCTTGAC AAGGGTGAGG 4400

CAGTTAACTT TTTGAAAGGA GCAGTTGTGA CAGCAGATCG AATTGTGACC 4450

GTCAGTCAGG TGAAATACTC AATACTTCTC TTTTTTCTTT GCGGGATGTT 4500

CTTCAGTTCA ATTGCCCTGT CTTTCACCCA ATTAAGAAAT GATTTAATCT 4550

TTTGTTTCTA GGGTTATTCA TGGGAGGTCA CAACTGCTGA AGGTGGACAG 4600

GGCCTCAATG AGCTCTTAAG CTCCCGAAAA AGTGTATTGA ATGGTAACTA 4650

TATTTGAATC CACTTATCTT CTTCTGAAAC ATATTTACAG AAATAGATGG 4700

ATGGGTTGCA AGAATAAATT CAGTTTGCTC TTTCGGTATG AAGGAATTGT 4750

AAATGGAATT GACATTAATG ATTGGAACCC CACCACAGAC AAGTGTCTCC 4800

CTCATCATTA TTCTGTCGAT GACCTCTCTG GAAAGGTGTG TGGATAGTAC 4850

CCTATATAAT AACATGTATA TCTGATCTAG TACTTTCTTT TTCTTTGCTA 4900

GTTTGCTTCC CATGATGTTC TCACTAACTA ATCCTATGTG GTTTGGCATA 4950

CTTGTCAGGC CAAATGTAAA GCTGAATTGC AGAAGGAGCT GGGTTTACCT 5000

GTAAGGGAGG ATGTTCCTCT GGTTAGATAC AAACCCCTAA GATATATATT 5050
TTTTAAATCC CTAAAAAAAA CTTGCCGATC ATCTCATTAG CTTGATTCAC 5100
AGATTGGCTT TATTGGAAGA CTGGATTACC AGAAAGGCAT TGATCTCATT 5150
AAAATGGCCA TTCCAGAGCT CATGAGGGAG GACGTGCAGT TTGTAAGTTC 5200
ATATTCTTTT TCTTGAGACT AGAGTATAAA TCAAACATGT AGGTGTGGGG 5250
TGGTATAATA CAGACATAAG TTCCAGCTAT TGCTTCCATG AGAATTTTAA 5300
TGCTATTCAG TAATATGCTA CTGCAAGTTT TGAAACAAAG TTGGAAGCAA 5350
TAAATATATG TGTAGCACTG ACCATGCAGT GCCACTATAG CTGGAATGTC 5400
CTGTAGTCTA TGTGATCTAA CACACTCAAC AACATGTTTT CGCATACAAA 5450
CACATGCGTG CGCGCAACAA ACATACTCTA CAATAAAATT GGCTTGGTGA 5500
ACTGCAGACA TGCTCTTATC TCCATTCCAA CATTTCTTGT TTCAACATTG 5550
GCTGAAGACT AAGAGAAGGG GGACCCAGGG TGATGTAGCC AACTAGATCC 5600
AGTAAGGAAG CTAGCCGAGC CTAGGAGGAT TCGCTTAGGT AGCTGGAACG 5650
TAGGGTCTCT GACAGGGAAG CTTCGGGAGC TAGTCGATGC AGTGGTGAGG 5700
AGAGGTGTTG ATATCCTTTG CGTCCAAGAA ACCAAATGTA GGGGACAGAA 5750
GGCGAAGGAG GTGGAGGATA CCGGCTTCAA GCTGTGGTAC ATGGGACGGC 5800
TGCAAACAGA AATGGCGTAG GCATCTTGAT CAACAAGAGC CTTAAGTATG 5850
GAGTGGTAGA CGTCAAGAGA CGTGGGGACC GGATTATCCT CGTCAAGCTG 5900
GTAGTTGGGG ACTTAGTTCT CAATGTTATC AGCGTGTATG CCCCGCAAGT 5950
AGGCCACAAT GAGAACGCCA AGAGGGAGTT CTGGGAAGGC CTGGAAGACA 6000
TGGTTAGGAG TGTACCGATT GGCGAGAAGC TCTTCATAGG AGGAGACCTC 6050
AATGGCCACG TGGGTACATC TAACATAGGT TTTGAAGGGG CACATGGGGG 6100
CTTTGGCTAT GGCATCAAGA ATCAAGAAGA AGATGTCTTA CGCTTTGCTC 6150
TAGCCTACGA CATGATTGTA GCTAACACCC TCTTTAGAAA GAGAGAATCA 6200
CATCTGGTGA CTTTTAGTAG TGGCCAACAC TAGCCAGATC GATTTCATCC 6250
TCTCGAGAAG AGAAGATAGG TGTGCGCGCC TAGACTGCAA GGTGATACCT 6300
TCGGATTCGT GTCCAGCGGG ATAAGCGTGC CAAAGTCGCT AGAATGAAGT 6350
GGTGGAAGCT CAAGGGGGAG GTAGCTCAGG CGTTCAAGGA GAGGGTCATT 6400

AGGGAGGGCC CTTGGGAGGA AGGAGGGGAT GCGGACAATG TGTGGATGAA 6450

GATGGCGACT TGCATTCGTA AGGTGGCCTC GGAGGAGTGT GGAGTGTCCA 6500

GGGGATGGAG AAGCGAAGAT AAGGATACCT GGTGGTGGAA TGATGATGTC 7000

CAGAAGGCAA TTAAAGAGAA GAAAGATTGC TTTAGACGCC TATACTTGGA 7050

TAGGAGTGCA GTCAACATAG AAAAGTACAA GATGGCGAAG AAGGCCGCAA 7100

AGCGAGCTGT CAGTGAAGCA AGGGGTCGGG CATATGAGGA TCTCTACCAA 7150

CGGTTAGGCA CGAAGGAAGG CGAAAGGGAC ATCTATAAGA TGGCCAAGAT 7200

CCGAGAGAGA GGAAGACGAG GGATATTGGC CAAGTCAAAT GCATCAAGGA 7250

TGGAGCAGAC CAACTCTTGG TGAAGGACGA GGAGATTAAG CATAGATGGC 7300

GGGAGTACTT CGACAAGCTG TTCAATGGGG AGGATGAGAG TCCTACCATT 7350

GAACTTGACG ACTCCTTTGA TGAGACCATC ATGCGTTTTA TGCGGCGAAT 7400

CCAGGAGTCC GAGGTCAAGG AGGCTTTAAA AAGGAGGCAA GGCGATGGGC 7450

CCTGATTGTA TCCCCATTGA GGTGTGGAAA GGCCTCGGGG ACATAGCGAT 7500

AGTATGGCTA ACCAAGCTAT TCAACCTCAT TTTTCGGGCA AACAAGATGC 7550

CAGAAGAATG GAGACGAAGT ATATTAGTAC CAATCATCAA ACAGGGGGGA 7600

TGTTCAGAGT TGTACTAATT ACCATGGAAT TAAGCTGATG AGCCATACAA 7650

TGAAGCTATG GGAGAGAATC ATTGAGCACC GCTTAAGAAG AATGACAAGC 7700

GTGACCAAAA ATCAGTTTGG TTTCATGCCT GGGAGGTCGA CCATGGAAAC 7750

CATTTTCTTG GTACGACAAC TTATGGAGAG ATACAGGGAG CAAAAGAAGG 7800

ACTTGCATAT GGTGTTCATT GACTTGAAGA AGGCCTATAA TAAGATACCG 7850

CGGAATGTCA TGTGGTGGGC CTTGGAGAAA CACAAAGTCC CAGCAAAGTA 7900

CATTACCCTC ATCAAGGACA TGTACGATAA TGTTGTGACA AGTGTTCGAA 7950

CAAGTGATGT CGACACTAAT GACTTCCCGA TTAAGATAGG ACTGCATCAG 8000

GGGTCAGCTT TGAGCCCTTA TCTTTTTGCC TTGGTGATGG ATGAGGTCAC 8050

AAGGGATATA CAAGGAGATA TCCCATGGTG TATGCTCTTT GTGGATGATT 8100

TGGTGCTAGT TGACGATAGT CGGGCGGGGG TAAATAACAA GTTAGAGTTA 8150

TGGAGACAAA CCTTGGAATC GAAAGGGTTT AGGCTTAGTA GAACTAAAAC 8200

CGAGTACATG ATGTGCGGTT TCAGTACTAC TAGGTGTGAG GAGGAGGAGG 8250

TTAGCCTTGA TGGGCAGGTG GTACCCCAGA AGGACACCTT TCGATATTTG 8300
GGGTCAATGC TGCAGGAGGA TGGGGGTATT GATGAAGATG TGAACCATCG 8350
AATCAAAGCT GGATGGATGA AGTGGCGCCA AGCTTCTGGC ATTCTTTGTG 8400
ACAAGAGAGT GCCACAAAAG CTAAGGCAAG TTCTACAGGA CGGCGGTTCG 8450
ACCCGCAATG TTGTATGGCG CTGAGTGTTG GCCGACTAAA AGGCGACATG 8500
TTCAACAGTT AGGTGTGGCG GAGATGCGTA TGTTGAGATG GATGTGTGGC 8550
CACACGAGGA AGRATCGAGT CCGGAATGAT GATATACGAG ATAGAGTTGG 8600
GGTAGCACCA ATTGAAGAGA AGCTTGTCCA ACATCGTCTG AGATGGTTTG 8650
GGCATATTCA GCGCACGCCT CCGAAAACTC CAGTGCATAA CGGACGGCTA 8700
AAGCGTGCGG AGAATGTCAA GAGAGGGCGG GGTAGACCGA ATTTGACATG 8750
GGAGGAGTCC GTTAAGAGAG ACCTGAAGGT TTGGAGTATT ACGAAAGAAC 8800
TAGCTATGGA CARGGGTGCG TGGAAGCTTG TTATCCATGT GCCAGAGCCA 8850
TGAGTTGATC ACGAGATCTT ATGGGTTTCA CCTCTAGCCT ACCCCAACTT 8900
GTTTGGGACT AAAGGCTTTG TTGTTGTTGT TGTTGTTGTT GTTGTAGCCA 8950
ACTAAATCCA GTTGATCAGT GGTTTTTACT CTTATTTTTA CAGGTCATGC 9000
TTGGATCTGG GGATCCAATT TTTGAAGGCT GGATGAGATC TACCGAGTCG 9050
AGTTACAAGG ATAAATTCCG TGGATGGGTT GGATTTAGTG TTCCAGTTTC 9100
CCACAGAATA ACTGCAGGGT ATGCCGAGAA CTTCTTAACA AGACCTTCGT 9150
TATCAGCTTG GATATATTAT AATGTTCAAA ACATTTATGT CTCTCTTTTT 9200
GTGCAGTTGC GATATATTGT TAATGCCATC CAGGTTTGAA CCTTGTGGTC 9250
TTAATCAGCT ATATGCTATG CAATATGGTA CAGTTCCTGT AGTTCATGGA 9300
ACTGGGGGCC TCCGAGTAAG ACAACTGCCT TGAAAATTAT CGTTATCTTG 9350
GCTCCAACGC AAATGTTCTA ATTGGCTCGT GTATTCAACA GGACACAGTC 9400
GAGACCTTCA ACCCTTTTGG TGCAAAAGGA GAGGAGGGTA CAGGGTACGC 9450
ACTGCTCAAT TTTAGCTAAC TTTCAGTTTA TCTTTTTGCA ATGTCTTGGG 9500
GGTTCATTGC GCCATAAATC AACTTGTGAT AATTAACTGT TACTGTTCTG 9550
TACTTGCAGG TGGGCGTTCT CACCGCTAAC CGTGGACAAG ATGTTGTGGG 9600
TAAGTTTTTG CTGAGCTCTT GTCCGGTTAT AGGATCGACC TTGGCTGTAG 9650

CATGGTACCT TAGTGCCCCT TGTATATAGA CCTAACCTGA TGGACTCACT 9700
TTGTCTACAC TAATCATAGT AGTCGATTGC CCGGAGGCGT TTTGCTTGGA 9750
TTCTGCTAAT TTAATTTTCA TGACGATAAC TCATACCATG GTTTGGTTCT 9800
CCGATGGGGG CCAGAATGGC GTCTAGTGTC TGCGATCTGT GTAACTAGCC 9850
AATGCCGGGT TGTTCCAAGT GAAAATTTAC CTTTTGACCA TTGTGCAGGC 9900
ATTGCGAACC GCGATGTCGA CATTCAGGGA GCACAAGCCG TCCTGGGAGG 9950
GGCTCATGAA GCGAGGCATG ACGAAAGACC ATACGTGGGA CCATGCCGCC 10000
GAGCAGTACG AGCAGATCTT CGAATGGGCC TTCGTGGACC AACCCTACGT 10050
CATGTAGACG GGGACTGGGG AGGTCGAAGC GCGGGTCTCC TTGAGCTCTG 10100
AAGACATGTT CCTCATCCTT CCGCGGCCCG GAAGGATACC CCTGTACATT 10150
GCGTTGTCCT GCTACAGTAG AGTCGCAATG CGCCTGCTTG CTTGGTCCGC 10200
CGGTTCGAGA GTAGATGACG GCTGTGCTGC TGCGGCGGTG ACAGCTTCGG 10250
GTGGATGACA GTTACAGTTT TGGGGAATAA GGAAGGGATG TGCTGCAGGA 10300
TGGTTAACAG CAAAGCACCA CTCAGATGGC AGCCTCTCTG TCCGTGTTAC 10350
AGCTGAAATC AGAAACCAAC TGGTGACTCT TTAGCCTTAG CGATTGTGAA 10400
GTTTGTTGCA TTCTGTGTAT GTTGTCTTGT CCTTAGCTGA CAAATATTTG 10450
ACCTGTTGGA TAATTCTATC TTTGCTGCTG TTTTTCTTTT GGTCAAAAGA 10500
GGGGTTCCCT CCGATTTCAT TAACGAAACC ACCAAAATAA CAGCACCCAG 10550
TGCAGGTCTC AGGTTCAGAT ATACTTAAGA CTACTAAATC TAACAGCAGC 10600
TAAAAAGCTT AAAGATTCAG GCGACATAAC CGAACAAAAT CCACAACCGA 10650
AGGGACCAAA GCAGGACAAG TAAAAAGGCA GNCGACACAA AGCGCAGGTC 10700
GCTGAAAAGG CAAGCAGACA GAGGTCTGCA TTCTGTCAAC ACCACTTGTG 10750
AAAAATGAAG AGAAGATCGA GAATTCCCGG GAATCCG 10787

(2) INFORMATION FOR SEQ ID NO: 14:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 647 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(ix) FEATURE:
(A) NAME/KEY: Protein
(B) LOCATION:1..647
(D) OTHER INFORMATION:/product= "deduced amino acid sequence for SSS I"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
Met Ala Ala Thr Gly Val Gly Ala Gly Cys Leu Ala Pro Ser Val Arg
1               5                   10                  15

Leu Arg Ala Asp Pro Ala Thr Ala Ala Arg Ala Ser Ala Cys Val Val
            20                  25                  30

Arg Ala Arg Leu Arg Arg Leu Ala Arg Gly Arg Tyr Val Ala Glu Leu
        35                  40                  45

Ser Arg Glu Gly Pro Ala Ala Arg Pro Ala Gln Gln Gln Gln Leu Ala
    50                  55                  60

Pro Pro Leu Val Pro Gly Phe Leu Ala Pro Pro Pro Pro Ala Pro Ala
65                  70                  75                  80

Gln Ser Pro Ala Pro Thr Gln Pro Pro Leu Pro Asp Ala Gly Val Gly
                85                  90                  95

Glu Leu Ala Pro Asp Leu Leu Leu Glu Gly Ile Ala Glu Asp Ser Ile
            100                 105                 110

Asp Ser Ile Ile Val Ala Ala Ser Glu Gln Asp Ser Glu Ile Met Asp
        115                 120                 125

Ala Asn Glu Gln Pro Gln Ala Lys Val Thr Arg Ser Ile Val Phe Val
    130                 135                 140

Thr Gly Glu Ala Ala Pro Tyr Ala Lys Ser Gly Gly Leu Gly Asp Val
145                 150                 155                 160

Cys Gly Ser Leu Pro Ile Ala Leu Ala Ala Arg Gly His Arg Val Met
                165                 170                 175

Val Val Met Pro Arg Tyr Leu Asn Gly Ser Ser Asp Lys Asn Tyr Ala
            180                 185                 190

Lys Ala Leu Tyr Thr Gly Lys His Ile Lys Ile Pro Cys Phe Gly Gly
        195                 200                 205

Ser His Glu Val Thr Phe Phe His Glu Tyr Arg Asp Asn Val Asp Trp
    210                 215                 220

Val Phe Val Asp His Pro Ser Tyr His Arg Pro Gly Ser Leu Tyr Gly
225                 230                 235                 240

Asp Asn Phe Gly Ala Phe Gly Asp Asn Gln Phe Arg Tyr Thr Leu Leu
                245                 250                 255

Cys Tyr Ala Ala Cys Glu Ala Pro Leu Ile Leu Glu Leu Gly Gly Tyr
            260                 265                 270
```

```
Ile Tyr Gly Gln Asn Cys Met Phe Val Val Asn Asp Trp His Ala Ser
        275                 280                 285

Leu Val Pro Val Leu Leu Ala Ala Lys Tyr Arg Pro Tyr Gly Val Tyr
    290                 295                 300

Arg Asp Ser Arg Ser Thr Leu Val Ile His Asn Leu Ala His Gln Gly
305                 310                 315                 320

Leu Glu Pro Ala Ser Thr Tyr Pro Asp Leu Gly Leu Pro Pro Glu Trp
                325                 330                 335

Tyr Gly Ala Leu Glu Trp Val Phe Pro Glu Trp Ala Arg Arg His Ala
            340                 345                 350

Leu Asp Lys Gly Glu Ala Val Asn Phe Leu Lys Gly Ala Val Val Thr
        355                 360                 365

Ala Asp Arg Ile Val Thr Val Ser Gln Gly Tyr Ser Trp Glu Val Thr
    370                 375                 380

Thr Ala Glu Gly Gly Gln Gly Leu Asn Glu Leu Leu Ser Ser Arg Lys
385                 390                 395                 400

Ser Val Leu Asn Gly Ile Val Asn Gly Ile Asp Ile Asn Asp Trp Asn
                405                 410                 415

Pro Thr Thr Asp Lys Cys Leu Pro His His Tyr Ser Val Asp Asp Leu
            420                 425                 430

Ser Gly Lys Ala Lys Cys Lys Ala Glu Leu Gln Lys Glu Leu Gly Leu
        435                 440                 445

Pro Val Arg Glu Asp Val Pro Leu Ile Gly Phe Ile Gly Arg Leu Asp
    450                 455                 460

Tyr Gln Lys Gly Ile Asp Leu Ile Lys Met Ala Ile Pro Glu Leu Met
465                 470                 475                 480

Arg Glu Asp Val Gln Phe Val Met Leu Gly Ser Gly Asp Pro Ile Phe
                485                 490                 495

Glu Gly Trp Met Arg Ser Thr Glu Ser Ser Tyr Lys Asp Lys Phe Arg
            500                 505                 510

Gly Trp Val Gly Phe Ser Val Pro Val Ser His Arg Ile Thr Ala Gly
        515                 520                 525

Cys Asp Ile Leu Leu Met Pro Ser Arg Phe Glu Pro Cys Gly Leu Asn
    530                 535                 540

Gln Leu Tyr Ala Met Gln Tyr Gly Thr Val Pro Val Val His Gly Thr
545                 550                 555                 560

Gly Gly Leu Arg Asp Thr Val Glu Thr Phe Asn Pro Phe Gly Ala Lys
                565                 570                 575

Gly Glu Glu Gly Thr Gly Trp Ala Phe Ser Pro Leu Thr Val Asp Lys
            580                 585                 590

Met Leu Trp Ala Leu Arg Thr Ala Met Ser Thr Phe Arg Glu His Lys
        595                 600                 605
```

```
Pro Ser Trp Glu Gly Leu Met Lys Arg Gly Met Thr Lys Asp His Thr
    610                 615                 620

Trp Asp His Ala Ala Glu Gln Tyr Glu Gln Ile Phe Glu Trp Ala Phe
625                 630                 635                 640

Val Asp Gln Pro Tyr Val Met
                645
```

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5072 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION:1..4993
(D) OTHER INFORMATION:/function= "region containing promoter of SSS I"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```
TCTAGATGCA TGCTGGATAG CGGTCGATGT GTGGAGTAAT AGTAGTAGAT GCAGAATCGT   60
TTCGGTCTAC TTGTCGCGGA CGTGATGCCT ATATACATGA TCATACCTAG ATATTCTCAT  120
AACTATGCTC AATTCTATCA ATTGCTCGAC AGTAATTCGT TTACCCACCG TAATACTTAT  180
GATCTTGAGA GAAGTCACTA GTGAAACCTA TGCCCCCCAG GTCTATTTTG CATCATATTA  240
ATCTTCCAAT ACTTAGTTAT TTCCATTGCC GTTTATTTTA CTTTGTATCT TTATTTCTTT  300
TTATTATAAA AAATACCAAA AATATTATCT TATCATATCT ATCAGATCTC ATTCTCGTAA  360
GTGACCGTGA AGGGATTGAC AACCCCTTTA TCGTGTTGGT TGCGAGGTTC TTGTTTGTTT  420
GTGTAGGTGC GTGTGACTCG CACGTCTCCT ACTGGATTGA TACCTTGGGT TTTCAAAAAC  480
TGAGAAAAAT ACTTACGCTA CTTTACTGCA TAACCCTTTC CTCTTTAAAA AAAAAAACCA  540
ACGTAGTATT CAAGAGGTAG CACGCTACCA TCCTCTCCAA CAGGAGCGCG GAGATCTTTG  600
TCCGGCAGGT TGATGCGGGC CGGGGAAGAA CTCCAGCTGC CTTGGCCAGC TTGGTCGTGA  660
GCCGCCCCAG CGGCGTCTTG AACCTGTCCA CGTAGCGCTC CCTGACACGC GGCGTGAACT  720
GAGAAGGCTT GTCGATGAAC TCCAGCTGTT GTGCCAGCCT AGCTTGCGCC TTCTTCTGCT  780
GGGTCATGCC CTTCGAGAAA CCCACCTTGG CCACCCTTGT GCTTGAGCGG CGCGCCACCT  840
CAGCAGGCGG CGGCGTGGGG ATGAAGAGGG TGTCTGCTTC CGGAGCAGGC GGGTCGGCGT  900
```

TGAACTTGAA AGGCGGTGGC CCCATGATGG ATGGGGGGAG CATGCCAAAG ACTTGGTTGA 960
GGAAAGTGGT GTTGGCGTCC ACCTCCAGTG CCTGCAGTTT GGAAGCCAGA CGATTGGCGT 1020
CGATCTCTGG CTCCGGCTGG AAGGAGGCTC GACGCTCCGG TGTGCCAGAA CGCAAAGGGA 1080
GGAGCGGCAG CTCTGGCTGA GCAGACCCCG CGCCCATGTA CTCTGCATTG GGCCAAGGCT 1140
GCAGGGGCAA GCCACCGGGA TGGGGGCGCG AGGTGGACTG CGCACCGGAG GAAGGCCAAG 1200
CTCAACCTCG GTGAGGTTCG CCCCAGACCA GGGCGGCAGG CTCGGGTCCA CAAAGGGCCA 1260
AACCGCCTCG TCCGCCCCGA AACTGTCCAG GACAGACGGC GGACGACGGA AGGCCGTGTC 1320
GTCGAGCTCG AGCAGCAGAG GGTCCGTGCG GGTGATGTCT TGCCAAATGG ACTCCACCTC 1380
CAGCAGGAAG GGGGACTGGT CCATCGCCCC TGGCCAAGCC ACTGGTACGC CAAAGATGGC 1440
ATCAGCAGCG TTTGCACCAG GGGGAGCAGC CACACCTTGG AGGACAGGGA GGGTGCGGAC 1500
GTCGACGGCA GCAAAACGTG GCTGGAGCAA GTTGCCGTCG CGTGCCGGCC TCGGCGAGCG 1560
CGAGCGGCTG TAGGAGCGCT CGGTGCCCTC AGACTCGGAC AGTGCGCCAG TGGGAGAGCC 1620
ATGGCGACGC CGGCCACCAC TGGACGTGCC ATGGCGCTGG TCCTGACGGC GCCTGGATGG 1680
CCCGTCCTCG CGGGCAGCTC CACCTGAGCG GCACCCGAGG AGCACACCCC GCCAAGCTGG 1740
GCCAGGGCGG CTGCGGCGAC GGCGACGGCC GCGGTCGCGG TCTGCACCAT CATCTTCATC 1800
TTCGTCATCG TGGCGCCTCG GACAAGGATG CTCGCTGTCA CCGACGCGAG GGACGTGAGC 1860
CGGCTCAGCC CGCCCTTCCT CGACGTGGCG AGCCCTGCGG ATATGCTCCT CGAGCGGCCA 1920
TTGGGGGTCG TTGGCGCGCG GCATCTCGGG GTCGCGGTCA GCTATCGGGG TGTAGTCCTT 1980
TGTGGTGTCC AGGTGGATGA GCAGAGAGAA ATCCGGCCCC TCTAGCCCCT CGTCCCGGGG 2040
GCAGCCCTCC GGCAGCGTCT GGCGGCCCCT GGGGTCCAGG GGTCGATCGA TGATGGAGAA 2100
CCCCCTTTTG GTGGGGATGT CGTCCGGACT CCATGCCCAC ACCCAGGCAA AGAGGCAGGC 2160
CGTGTTGGAG AGGGAGGTCG TCTGCCGCTC CAACCAGTCG ACGTGGCATG TCTTCCCGAG 2220
CGCATCCTGC CCCGCCTCCT TGTTCCAGGA CTGCACCGGC ATGTTCTCGA CGGCGATGCG 2280
GCAGTAGTAC CGCCAGACAC GGCGGTGGCC GTGTGCCGAT GGTGACCAGG CCGACAGGGA 2340
GAGCGCGACG CCCCAGCAGG AGACGACCCC AGCGTCGAAA GCGATGTCCC GGTGCCTGAA 2400
GTGGACGAGC CCAGAGATGG CCAGGCGCAT TGACGCGGGG AAGGGGAAGG AGTTAGGATG 2460
GGCGACGCGG CCGGAGTGAA CCGCGGCGTG GTGGCCGACG GGGCTGGAGA GGCAGAGGCG 2520
GAGTCATCCG AGAGAGGTGT ATCAGTGGCT CTGCACAATA CCCAGTGTCG CCACATCATA 2580
TCCTGCTGAA TAACCACACA TGTGTACTGT CGTTAAATAA ATCATTGGTC ACGCGAACCC 2640
GGAAAAAGAC GGCGAAAAAT TCACGGACAC ACGACTAGTA GTACCCAATA TACTCGGCAA 2700
AAACAGTGAC ACGTCGTTTT GCGTTGTCGG CCGGTGTTGT CGAGTCATTG TACTATGTTT 2760
TGTCGTTTCT TTCTTTTCTC CAAATCGACA AACCGTTTGT CTTTGGTTAA AAAACAGAAA 2820
CATACAAAAT CAAATGAATG CATTCAAGGG CCGGTAATCC AATTCTGAGC CCAGGCTCAG 2880
CTACACCCGC CCTTACAAAA AAATCAAAAT AAATACTAGA AAAATTCAAA AAATTCCAAT 2940

TTGTTTGTGC GTGGTAGATA ATTTGATGCG TGAGGTACGC TTCAATTTTC AAATTATTTG 3000
GACATCTGAG CAGCTCTCAG CAAAAAAGAC AAATTCGGGG TCTGTAAAAA TGTTTACTGT 3060
TCATGCACTG TTCTGACCCG ATTTGTCTTT TTTGCTGAGA GCTTCTCAGA AGTCCAAATG 3120
AGCTAAAATT TTGAGCGGAG CTTACGTGAT AAAATGTCTA TCATGCAAAA AAGGATTGGA 3180
ATTTTTTGAA TTTTTTTTAT TTTTTGTGAT TTGTTTCCTG GACGGGTGCA GATAAGCCTG 3240
GGCACCGAAA CGCCGCACTC AGGCTCATCC TTTTCTATAA AAGAAAAGAA ATACATACAA 3300
TTTCCCTCTG TTTTTTGAGC AAGGGGCACC ACCCACCAAA GAGTTTTCAA CTCACATGGT 3360
ATTAGAGCAT CTACAGCCGG GCGTCTCAAA CCAGCCTCAT ACGCTTGAGC GGGTCGCCTT 3420
GGTCACGATT TTTTGACCCA GACGGGCCCC TCAAACGGTC CTTAAACGCC CAGGCTGACC 3480
GACAACCCAC ATATCCAGCC CAAATATGGG GTGGATATGG GGGCGCCCGG GCACGCCAGC 3540
CCGCGGACAC CACACATCTT CAGTTTCTAA TTTGAGATAT CCGGATGTGG AATGCGTTTT 3600
TGAGGGGTGA CCGGTCCCTG TCCGTGGATG CGCCCGGACG TTTGAGGGGT TGGATTTGCC 3660
AAGTCTGATT AGAGATGCTC TTAGGTGTTC CACCCCCATC CCTTGATGGC TAGGGCAAAC 3720
TCTCCCCTCC AAACTTTGTC GGCGAGCCTG TGGATTCTTC TCTCCTCTGC CCGCTGCTCC 3780
GGCGGCTGAT GGCGGGGAGG AGAATCCCGG TGTCTTCGCT TGGTTAGTTG TTTAAGTTAC 3840
GTACTTTTTT AGTCCTCGCA GGTGCGGCGT TCGGACGTAT GGTCGTGCTT CTTTTTTGAG 3900
TTTGTCTTCC GGGCTCTGAT CCTCCTCGAG TTCGTCCATC TGGACGTACT CGACGGAGCT 3960
CCGGCATAGA TTCCTATCAT CGTCTTGGTG AGGTGAGGTT ATGGTTTCTT GTCATGTGGG 4020
CAGATTTGGT GCCAGATGCT TCATATCTAT TCAAGGGTTC AGCGGCAACA ACTGCGGCTC 4080
CAGAGCGATG GTCCTTAAGG GCACGTGCAC GAAGACTTCA CGGCTGTTAT CGACAAGGTC 4140
AAGCCGGCTC CGATAGGGGA GCAGCGACAG CGGCGCGTCA ACCGCTCGTT CTGGCGGCAG 4200
TAGTGGTCGT TCGGTGCTCT CGGAACCTCG ATGTAATTTT TATGATTTTA GAGATGCTTT 4260
GTACTTCCGA TCGATGAACT CTGATAATAG ATATCTCTTC TCTCGCAAAA AAAGAGAGTT 4320
TTCAACTGAA AACAAAAGAG TTTCACTAGT TCTTCTTTTA GAAACAGAGT TTCACTAGCA 4380
CTTTTTTTTG CGAGAAGTCG AGTTTCACTA AGTACTAAAC CCACGCAATT ATTCTCAAAA 4440
AAAAAACCCA CGCAACTGTC TGGATCCATC TTCGTTTTTT CCCCGAGAAT CGTCTGGATC 4500
CATTTTCGTG TGCGAGGCAT CCTCTCATTT TGCACGGCCC AGCTCTCTTC TCGCCGGCGT 4560
ACGCTGCTAC ATGTCGGCAC TCCACGCAAA CAAAAAGAAG CCCAACCGAA AACGCACGCG 4620
CCTTTCCAGG CTCACCACGG AAAAAAATAC CACGCGCCGC TCACGAGCAA ACCGTGACAA 4680
CAGCCAGCCA GATATGGCAA CGGAGGCACG GGCCGCACAC AGCCACTGAA AACCGCAGCT 4740
GCTCTTCCGT CCGTCCGTCC CTCCGCCCGT CCGCGCCACT CCACTCGCCT TGCCCCACTC 4800
CCACTCTTCT CTCCCCGCGC ACACCGAGTC GGCACCGGCT CATCACCCAT CACCTCGGCC 4860
TCGGCCACCG GCAAACCCCC CGATCCGCTT TTGCAGGCAG CGCACTAAAA CCCCGGGGAG 4920

CGCGCCCCGC GGCAGCAGCA GCACCGCAGT GGGAGAGAGA GGCTTCGCCC CGGCCCGCAC 4980

CGAGCGGGGC GATCCACCGT CCGTGCGTCC GCACCTCCTC CGCCTCCTCC CCTGTCCCGC 5040

GCGCCCACAC CCATGGCGGC GACGGGCGTC GG 5072

(2) INFORMATION FOR SEQ ID NO: 16:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1706 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..1706
(D) OTHER INFORMATION:/product= "partial cDNA for hexaploid wheat DBE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

```
GCT GTG TCG AAG CTT GAC TAT TTG AAG GAG CTT GGA GTT AAT TGT ATT      48
Ala Val Ser Lys Leu Asp Tyr Leu Lys Glu Leu Gly Val Asn Cys Ile
  1               5                  10                  15

GAA TTA ATG CCC TGC CAT GAG TTC AAC GAG CTG GAG TAC TCA ACC TCT      96
Glu Leu Met Pro Cys His Glu Phe Asn Glu Leu Glu Tyr Ser Thr Ser
             20                  25                  30

TCT TCC AAG ATG AAC TTT TGG GGA TAT TCT ACC ATA AAC TTC TTT TCA     144
Ser Ser Lys Met Asn Phe Trp Gly Tyr Ser Thr Ile Asn Phe Phe Ser
         35                  40                  45

CCA ATG ACG AGA TAC ACA TCA GGC GGG ATA AAA AAC TGT GGG CGT GAT     192
Pro Met Thr Arg Tyr Thr Ser Gly Gly Ile Lys Asn Cys Gly Arg Asp
     50                  55                  60

GCC ATA AAT GAG TTC AAA ACT TTT GTA AGA GAG GCT CAC AAA CGG GGA     240
Ala Ile Asn Glu Phe Lys Thr Phe Val Arg Glu Ala His Lys Arg Gly
 65                  70                  75                  80

ATT GAG GTG ATC CTG GAT GTT GTC TTC AAC CAT ACA GCT GAG GGT AAT     288
Ile Glu Val Ile Leu Asp Val Val Phe Asn His Thr Ala Glu Gly Asn
                 85                  90                  95

GAG AAT GGT CCA ATA TTA TCA TTT AGG GGG GTC GAT AAT ACT ACA TAC     336
Glu Asn Gly Pro Ile Leu Ser Phe Arg Gly Val Asp Asn Thr Thr Tyr
            100                 105                 110

TAT ATG CTT GCA CCC AAG GGA GAG TTT TAT AAC TAT TCT GGC TGT GGG     384
Tyr Met Leu Ala Pro Lys Gly Glu Phe Tyr Asn Tyr Ser Gly Cys Gly
        115                 120                 125
```

AAT ACC TTC AAC TGT AAT CAT CCT GTG GTT CGT CAA TTC ATT GTA GAT 432
Asn Thr Phe Asn Cys Asn His Pro Val Val Arg Gln Phe Ile Val Asp
130 135 140

TGT TTA AGA TAC TGG GTG ATG GAA ATG CAT GTT GAT GGT TTT CGT TTT 480
Cys Leu Arg Tyr Trp Val Met Glu Met His Val Asp Gly Phe Arg Phe
145 150 155 160

GAT CTT GCA TCC ATA ATG ACC AGA GGT TCC AGT CTG TGG GAT CCA GTT 528
Asp Leu Ala Ser Ile Met Thr Arg Gly Ser Ser Leu Trp Asp Pro Val
165 170 175

AAC GTG TAT GGA GCT CCA ATA GAA GGT GAC ATG ATC ACA ACA GGG ACA 576
Asn Val Tyr Gly Ala Pro Ile Glu Gly Asp Met Ile Thr Thr Gly Thr
180 185 190

CCT CTT GTT ACT CCA CCA CTT ATT GAC ATG ATC AGC AAT GAC CCA ATT 624
Pro Leu Val Thr Pro Pro Leu Ile Asp Met Ile Ser Asn Asp Pro Ile
195 200 205

CTT GGA GGC GTC AAG CTC ATT GCT GAA GCA TGG GAT GCA GGA GGC CTC 672
Leu Gly Gly Val Lys Leu Ile Ala Glu Ala Trp Asp Ala Gly Gly Leu
210 215 220

TAT CAA GTA GGT CAA TTC CCT CAC TGG AAT GTT TGG TCT GAG TGG AAT 720
Tyr Gln Val Gly Gln Phe Pro His Trp Asn Val Trp Ser Glu Trp Asn
225 230 235 240

GGG AAG TAC CGG GAC ATT GTG CGC CAA TTC ATT AAA GGC ACT GAT GGA 768
Gly Lys Tyr Arg Asp Ile Val Arg Gln Phe Ile Lys Gly Thr Asp Gly
245 250 255

TTT GCT GGT GGT TTT GCC GAA TGT CTT TGT GGA AGT CCA CAC CTA TAC 816
Phe Ala Gly Gly Phe Ala Glu Cys Leu Cys Gly Ser Pro His Leu Tyr
260 265 270

CAG GCA GGA GGA AGG AAA CCT TGG CAC AGT ATC AAC TTT GTA TGT GCA 864
Gln Ala Gly Gly Arg Lys Pro Trp His Ser Ile Asn Phe Val Cys Ala
275 280 285

CAT GAT GGA TTT ACA CTG GGT GAT TTG GTA ACA TAT AAT AAC AAG TAC 912
His Asp Gly Phe Thr Leu Gly Asp Leu Val Thr Tyr Asn Asn Lys Tyr
290 295 300

AAT TTA CCA AAT GGG GAG AAC AAT AGA GAT GGA GAA AAT CAC AAT CTT 960
Asn Leu Pro Asn Gly Glu Asn Asn Arg Asp Gly Glu Asn His Asn Leu
305 310 315 320

AGC TGG AAT TGT GGG GAG GAA GGA GAA TTC GCA AGA TTG TCT GTC AAA 1008
Ser Trp Asn Cys Gly Glu Glu Gly Glu Phe Ala Arg Leu Ser Val Lys
325 330 335

AGA TTG AGG AAG AGG CAG ATG CGC AAT TTC TTT GTT TGT CTC ATG GTT 1056
Arg Leu Arg Lys Arg Gln Met Arg Asn Phe Phe Val Cys Leu Met Val
340 345 350

TCT CAA GGA GTT CCA ATG TTT TAC ATG GGC GAT GAA TAT GGC CAC ACA 1104
Ser Gln Gly Val Pro Met Phe Tyr Met Gly Asp Glu Tyr Gly His Thr
355 360 365

AAA GGG GGC AAC AAC AAT ACA TAC TGC CAT GAT TCT TAT GTC AAT TAT 1152
Lys Gly Gly Asn Asn Asn Thr Tyr Cys His Asp Ser Tyr Val Asn Tyr
370 375 380

```
TTT CGC TGG GAT AAA AAA GAA CAA TAC TCT GAC TTG CAC AGA TTC TGC   1200
Phe Arg Trp Asp Lys Lys Glu Gln Tyr Ser Asp Leu His Arg Phe Cys
385                 390                 395                 400

TGC CTC ATG ACC AAA TTC CGC AAG GAG TGC GAG GGT CTT GGC CTT GAG   1248
Cys Leu Met Thr Lys Phe Arg Lys Glu Cys Glu Gly Leu Gly Leu Glu
                405                 410                 415

GAC TTT CCA ACG GCC GAA CGG CTG CAG TGG CAT GGT CAT CAG CCT GGG   1296
Asp Phe Pro Thr Ala Glu Arg Leu Gln Trp His Gly His Gln Pro Gly
            420                 425                 430

AAG CCT GAT TGG TCT GAG AAT AGC CGA TTC GTT GCC TTT TCC ATG AAA   1344
Lys Pro Asp Trp Ser Glu Asn Ser Arg Phe Val Ala Phe Ser Met Lys
        435                 440                 445

GAT GAA AGA CAG GGC GAG ATC TAT GTG GCC TTC AAC ACC AGC CAC TTA   1392
Asp Glu Arg Gln Gly Glu Ile Tyr Val Ala Phe Asn Thr Ser His Leu
    450                 455                 460

CCG GCC GTT GTT GAG CTC CCA GAG CGC GCA GGG CGC CGG TGG GAA CCG   1440
Pro Ala Val Val Glu Leu Pro Glu Arg Ala Gly Arg Arg Trp Glu Pro
465                 470                 475                 480

GTG GTG GAC ACA GGC AAG CCA GCA CCA TAT GAC TTC CTC ACC GAC GAC   1488
Val Val Asp Thr Gly Lys Pro Ala Pro Tyr Asp Phe Leu Thr Asp Asp
                485                 490                 495

TTA CCT GAT CGC GCT CTC ACC ATA CAC CAG TTC TCT CAT TTC CTC AAC   1536
Leu Pro Asp Arg Ala Leu Thr Ile His Gln Phe Ser His Phe Leu Asn
            500                 505                 510

TCC AAC CTC TAC CCC ATG CTC AGC TAC TCA TCG GTC ATC CTA GTA TTG   1584
Ser Asn Leu Tyr Pro Met Leu Ser Tyr Ser Ser Val Ile Leu Val Leu
        515                 520                 525

CGC CCT GAT GTT TGA GAG ACA AAT ATA TAC AGT AAA TAA TAT GTC TAT   1632
Arg Pro Asp Val  *  Glu Thr Asn Ile Tyr Ser Lys  *  Tyr Val Tyr
    530                 535                 540

ATG TAG TCC TTT GGC GTA TTA TCA GTG TGC ACA ATT GCT CTA TTG CCA   1680
Met  *  Ser Phe Gly Val Leu Ser Val Cys Thr Ile Ala Leu Leu Pro
545                 550                 555                 560

GTG ATC TAT TCG ATA GCG GCC GCG  AA                               1706
Val Ile Tyr Ser Ile Ala Ala Ala
                565
```

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9289 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: triticum tauschii
(F) TISSUE TYPE: Endosperm

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..9289
(D) OTHER INFORMATION:/product= "genomic sequence of DBE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```
CGG GAC CGT CCC TTG GCA ACT TGG GTT ACG TTG GGA CCT GAC GCT TCG      48
Arg Asp Arg Pro Leu Ala Thr Trp Val Thr Leu Gly Pro Asp Ala Ser
    570                 575                 580

CTT ATC CGG TGT GCC CTG AGA CGA GAT ATG TGC AGC TCC TAT CGG ATT      96
Leu Ile Arg Cys Ala Leu Arg Arg Asp Met Cys Ser Ser Tyr Arg Ile
585                 590                 595                 600

TGT CGG CAC ATT CGG CGG CTT TGC TGG TCT TGT TTT ACC ATT GTC GAA     144
Cys Arg His Ile Arg Arg Leu Cys Trp Ser Cys Phe Thr Ile Val Glu
                605                 610                 615

ATG TCT TAT AAA CCG GGA TTC CGA GAC TGA TCG GGT CTT CCC GGG AGA     192
Met Ser Tyr Lys Pro Gly Phe Arg Asp  *  Ser Gly Leu Pro Gly Arg
            620                 625                 630

AGG TTT ATC CTT CGT TGA CCG TGA GAG CTT ATA ATG GGC TAA GTT GGG     240
Arg Phe Ile Leu Arg  *  Pro  *  Glu Leu Ile Met Gly  *  Val Gly
        635                 640                 645

ACA CCC CTG CAG GGT ATT ATC TTT CGA AAG CCG TGC CCG CGG TTA TGA     288
Thr Pro Leu Gln Gly Ile Ile Phe Arg Lys Pro Cys Pro Arg Leu  *
    650                 655                 660

GGC AGA TGG GAA TTT GTT AAT GTC CGA TTG TAG AGA ACC TGT CAC TTG     336
Gly Arg Trp Glu Phe Val Asn Val Arg Leu  *  Arg Thr Cys His Leu
665                 670                 675                 680

ACT TAA TTT AAA ATT CAT CAA CCG TGT GTG TAG CCG TGA TGG TCT CTT     384
Thr  *  Phe Lys Ile His Gln Pro Cys Val  *  Pro  *  Trp Ser Leu
            685                 690                 695

TTC GGC GGA GTC CGG GAA GTG AAC ACG GTT TGA GTT ATG CAT GAA CGT     432
Phe Gly Gly Val Arg Glu Val Asn Thr Val  *  Val Met His Glu Arg
            700                 705                 710

AAG TAG TTT CAG GAT CAC TCC TTG ATC ACT TCT AGC TCC GCG ACC GTT     480
Lys  *  Phe Gln Asp His Ser Leu Ile Thr Ser Ser Ser Ala Thr Val
        715                 720                 725

GCG TTG TTT CTC TTC TCG CTC TCA TTT GCG TAT GTT AGC CAC CAT ATA     528
Ala Leu Phe Leu Phe Ser Leu Ser Phe Ala Tyr Val Ser His His Ile
    730                 735                 740

TGC TTA GTG TCT GCT GCA GCT CCA CCT CAT TAC CCC TTC CTT TCC TAT     576
Cys Leu Val Ser Ala Ala Ala Pro Pro His Tyr Pro Phe Leu Ser Tyr
745                 750                 755                 760

AAG CTT AAA TAG TCT TGA TCT CGC GGG TGT GAG ATT GCT GAG TCC TCG     624
Lys Leu Lys  *  Ser  *  Ser Arg Gly Cys Glu Ile Ala Glu Ser Ser
                765                 770                 775
```

```
TGA CTT ACA GAT TCT ACC AAA ACA GTT GCA GGT GTC GAC GAT GCC AGT    672
 *  Leu Thr Asp Ser Thr Lys Thr Val Ala Gly Val Asp Asp Ala Ser
            780                 785                 790

GCA GGT GAC GCA ACC GAG CTC AAG TGG GAG TTC GAC GAG GAA CGT GGT    720
Ala Gly Asp Ala Thr Glu Leu Lys Trp Glu Phe Asp Glu Glu Arg Gly
        795                 800                 805

CGT TAC TAT GTT TCT TTT CCT GAT GAT CAG TAG TGG AGC CCA GTT GGG    768
Arg Tyr Tyr Val Ser Phe Pro Asp Asp Gln  *  Trp Ser Pro Val Gly
    810                 815                 820

ACG ATC GGG GAT CTA GCA TTT GGG GTT ATC TTA ATT TCT TTT AGA TTT    816
Thr Ile Gly Asp Leu Ala Phe Gly Val Ile Leu Ile Ser Phe Arg Phe
825                 830                 835                 840

GAC CGT AAT CGG TCT ATG TGT GGA TTT TGG ATG ATG TAT GAA TTA TTT    864
Asp Arg Asn Arg Ser Met Cys Gly Phe Trp Met Met Tyr Glu Leu Phe
                845                 850                 855

ATG TAT TGT GTG AAG TGG CGA TTG TAA GCC AAC TCT CGT TAT CCC ATT    912
Met Tyr Cys Val Lys Trp Arg Leu  *  Ala Asn Ser Arg Tyr Pro Ile
            860                 865                 870

CTT GTT CAT TAC ATG GGA TTG TGT GAA GAT GAC CCT TCT TGC GAC AAA    960
Leu Val His Tyr Met Gly Leu Cys Glu Asp Asp Pro Ser Cys Asp Lys
        875                 880                 885

ACC ACA ATG CGG TTA TGC CTC TAA GTC GTG CCT CGA CAC GTG GGA GAT   1008
Thr Thr Met Arg Leu Cys Leu  *  Val Val Pro Arg His Val Gly Asp
    890                 895                 900

ATA GCC GCA TCG TGG GCG TTA CAC GCA AGT CTT CAT AGC AAC CAA AAC   1056
Ile Ala Ala Ser Trp Ala Leu His Ala Ser Leu His Ser Asn Gln Asn
905                 910                 915                 920

TCC TCT CCG CAT TAC AAG CCA CCA ATC GCA GCC ACC ATG ACT TTC TTC   1104
Ser Ser Pro His Tyr Lys Pro Pro Ile Ala Ala Thr Met Thr Phe Phe
                925                 930                 935

ACC ACT GTC AAT GCC ATG AAA ATC TAT ATG TAG ACA TGT CCC ATT GCA   1152
Thr Thr Val Asn Ala Met Lys Ile Tyr Met  *  Thr Cys Pro Ile Ala
            940                 945                 950

TCG GCA AGA AAG CGA AGC TTC ACG GCA CAC CTT CAT GAA GCC TCT CTG   1200
Ser Ala Arg Lys Arg Ser Phe Thr Ala His Leu His Glu Ala Ser Leu
        955                 960                 965

GCC GAA GAC AAG GAT GCG CCC GAC CGG ATC AAT TCC TAT CTA GAT ACC   1248
Ala Glu Asp Lys Asp Ala Pro Asp Arg Ile Asn Ser Tyr Leu Asp Thr
    970                 975                 980

TAG TGG AGC CAT GCG CCA ATA GCG GAG ATC TCC GAG AGG AAG ACC GGA   1296
 *  Trp Ser His Ala Pro Ile Ala Glu Ile Ser Glu Arg Lys Thr Gly
985                 990                 995                1000

ACT CGT CGG ACG TCG GCG TCC AAA TCG AGG AGG CCG GCA TGA AGC ACA   1344
Thr Arg Arg Thr Ser Ala Ser Lys Ser Arg Arg Pro Ala  *  Ser Thr
                1005                1010                1015

TCG AGG ATG GTG ATC CCC ATA CGG GTA GAT CGG GTC GGC CGC CAT CTC   1392
Ser Arg Met Val Ile Pro Ile Arg Val Asp Arg Val Gly Arg His Leu
            1020                1025                1030
```

ACA CCG AGA TTA GGA TGC TTA AAA CGG TTT TTT TGG CAC TAG CAT TAT 1440
Thr Pro Arg Leu Gly Cys Leu Lys Arg Phe Phe Trp His * His Tyr
1035 1040 1045

TTT GCA TCA TCC GTT GGA GAG AAC ATG AGA GAG CCC CAT TTC TTC CAC 1488
Phe Ala Ser Ser Val Gly Glu Asn Met Arg Glu Pro His Phe Phe His
1050 1055 1060

GGT TCT ACC TAT GGG ATC TTG TTC TGC TTG CAA CCG GGC CTC ACG GAA 1536
Gly Ser Thr Tyr Gly Ile Leu Phe Cys Leu Gln Pro Gly Leu Thr Glu
1065 1070 1075 1080

AAC CCG CGC CAG CGG ACC CAC CCC ATG CTA GCA GGG CAC GGC ACC CGC 1584
Asn Pro Arg Gln Arg Thr His Pro Met Leu Ala Gly His Gly Thr Arg
1085 1090 1095

AGC GGC CGG TCC AAA TGG ACG GTG AGA ACC GCA ACG CGA CAC GCC CGG 1632
Ser Gly Arg Ser Lys Trp Thr Val Arg Thr Ala Thr Arg His Ala Arg
1100 1105 1110

CAC TGT CAG CAA AGC GAG AGC GCG CGC ACG GCA CAC GCA CGC TCG GAC 1680
His Cys Gln Gln Ser Glu Ser Ala Arg Thr Ala His Ala Arg Ser Asp
1115 1120 1125

GAA CGG ACG GTG CGA TCG ATC CCT CCC CCC TCG CTC AAC CAC AGT AGT 1728
Glu Arg Thr Val Arg Ser Ile Pro Pro Pro Ser Leu Asn His Ser Ser
1130 1135 1140

ACC CTG CCA CAC TAT CAC GCA CGC ACT CGA GTC ACA CCT CCC ACG AAG 1776
Thr Leu Pro His Tyr His Ala Arg Thr Arg Val Thr Pro Pro Thr Lys
1145 1150 1155 1160

AAC CAA CAG GAG GCG CGG ATC CCA CCG ATA AAT AAC CCC GCC TCG CCG 1824
Asn Gln Gln Glu Ala Arg Ile Pro Pro Ile Asn Asn Pro Ala Ser Pro
1165 1170 1175

CTC CTC CCC AAA ATC AAT CAC CGA TCG CTC GGG GTT CCC GGC ATG ACG 1872
Leu Leu Pro Lys Ile Asn His Arg Ser Leu Gly Val Pro Gly Met Thr
1180 1185 1190

ATG ATG GCC ATG GCC AAG GCG CCC TGC CTC TGC GCG CGC CCG TCC CTC 1920
Met Met Ala Met Ala Lys Ala Pro Cys Leu Cys Ala Arg Pro Ser Leu
1195 1200 1205

GCC GCG CGC GCG AGG CGG CCG GGG CCG GGG CCG GCG CCG CGC CTG CGA 1968
Ala Ala Arg Ala Arg Arg Pro Gly Pro Gly Pro Ala Pro Arg Leu Arg
1210 1215 1220

CGG TGG CGA CCC AAT GCG ACG GCG GGG AAG GGG GTC GGC GAG GTG TGC 2016
Arg Trp Arg Pro Asn Ala Thr Ala Gly Lys Gly Val Gly Glu Val Cys
1225 1230 1235 1240

GCC GCG GTT GTC GAG GCG GCG ACG AAG GCC GAG GAT GAG GAC GAC GAC 2064
Ala Ala Val Val Glu Ala Ala Thr Lys Ala Glu Asp Glu Asp Asp Asp
1245 1250 1255

GAG GAG GAG GCG GTG GCG GAG GAC AGG TAC GCG CTC GGC GGC GCG TGC 2112
Glu Glu Glu Ala Val Ala Glu Asp Arg Tyr Ala Leu Gly Gly Ala Cys
1260 1265 1270

AGG GTG CTC GCC GGA ATG CCC GCG CCG CTG GGC GCC ACC GCG CTC GCC 2160
Arg Val Leu Ala Gly Met Pro Ala Pro Leu Gly Ala Thr Ala Leu Ala
1275 1280 1285

GGC GGG GTC AAT TTC GCC GTC TAC TCC GGT GGA GCC ACC GCC GCG GCG 2208
Gly Gly Val Asn Phe Ala Val Tyr Ser Gly Gly Ala Thr Ala Ala Ala
1290 1295 1300

CTC TGC CTC TTC ACG CCA GAA GAT CTC AAG GCG GTG GGG TTG CCT CCC 2256
Leu Cys Leu Phe Thr Pro Glu Asp Leu Lys Ala Val Gly Leu Pro Pro
1305 1310 1315 1320

GAG TAG AGT TCA TCA GCT TTG CGT GCG CCG CGC GCC CCC TTT TCT GGC 2304
Glu * Ser Ser Ser Ala Leu Arg Ala Pro Arg Ala Pro Phe Ser Gly
1325 1330 1335

CTG CGA TTT AAG TTT TGT ACT GGG GGA AAT GCT GCA GGA TAG GGT GAC 2352
Leu Arg Phe Lys Phe Cys Thr Gly Gly Asn Ala Ala Gly * Gly Asp
1340 1345 1350

GGA GGA GGT TTC CCT TGA CCC CCT GAT GAA TCG GAC TGG GAA CGT GTG 2400
Gly Gly Gly Phe Pro * Pro Pro Asp Glu Ser Asp Trp Glu Arg Val
1355 1360 1365

GCA TGT CTT CAT TGA AGG CGA GCT GCA CGA CAT GCT TTA CGG GTA CAG 2448
Ala Cys Leu His * Arg Arg Ala Ala Arg His Ala Leu Arg Val Gln
1370 1375 1380

GTT CGA CGG CAC CTT TGC TCC TCA CTG CGG GCA CTA CCT TGA TAT TTC 2496
Val Arg Arg His Leu Cys Ser Ser Leu Arg Ala Leu Pro * Tyr Phe
1385 1390 1395 1400

CAA TGT CGT GGT GGA TCC TTA TGC TAA GGT GAT CAT ACT TTA GCT TTA 2544
Gln Cys Arg Gly Gly Ser Leu Cys * Gly Asp His Thr Leu Ala Leu
1405 1410 1415

CCT GCA TCT TGG TAT TTA CAG TAG AAA TTG TTA CGT GGA CCC TTA TTT 2592
Pro Ala Ser Trp Tyr Leu Gln * Lys Leu Leu Arg Gly Pro Leu Phe
1420 1425 1430

GTT GCC TTT TGT GTT GCT CTA GGC AGT GAT AAG CCG AGG GGA GTA TGG 2640
Val Ala Phe Cys Val Ala Leu Gly Ser Asp Lys Pro Arg Gly Val Trp
1435 1440 1445

CGT TCC GGC GCG TGG TAA CAA TTG CTG GCC TCA GAT GGC TGG CAT GAT 2688
Arg Ser Gly Ala Trp * Gln Leu Leu Ala Ser Asp Gly Trp His Asp
1450 1455 1460

CCC TCT TCC ATA TAG CAC GGT ATG CCT GAT TGC TGA AAA TAT TGG CTG 2736
Pro Ser Ser Ile * His Gly Met Pro Asp Cys * Lys Tyr Trp Leu
1465 1470 1475 1480

CAT TTG TTT CTC TCT TTT TCT CAT ATT TTT CTC CTG TCT TTC ACT TGT 2784
His Leu Phe Leu Ser Phe Ser His Ile Phe Leu Leu Ser Phe Thr Cys
1485 1490 1495

ACT ACA TTG CCT CAG ACA GTC ATG ATC AAA GAG AGC AGT GTC ATT AGA 2832
Thr Thr Leu Pro Gln Thr Val Met Ile Lys Glu Ser Ser Val Ile Arg
1500 1505 1510

CAT TTG TAG TTG TCT GCT GAC TTT GAC CAA AAC TTG TAA TTT ACT GTT 2880
His Leu * Leu Ser Ala Asp Phe Asp Gln Asn Leu * Phe Thr Val
1515 1520 1525

GTT AAA GGT CCT TGA ATC ATA TTT TTT TAT AAT ATT ATG TTT GCA AGT 2928
Val Lys Gly Pro * Ile Ile Phe Phe Tyr Asn Ile Met Phe Ala Ser
1530 1535 1540

```
GGA AGT AAA GTG AAA TTG CAT CTA GTA TTT GTT GTT GCT GTC TTA GTC   2976
Gly Ser Lys Val Lys Leu His Leu Val Phe Val Val Ala Val Leu Val
1545                1550                1555                1560

GTT TAA TTG GAC ATG CAG TAA AAA GGT TTG CAT CTG CAG TTT GAT TGG   3024
Val  *  Leu Asp Met Gln  *  Lys Gly Leu His Leu Gln Phe Asp Trp
                1565                1570                1575

GAA GGC GAC CTA CCT CTA AGA TAT CCT CAA AAG GAC CTG GTA ATA TAT   3072
Glu Gly Asp Leu Pro Leu Arg Tyr Pro Gln Lys Asp Leu Val Ile Tyr
            1580                1585                1590

GAG ATG CAC TTG CGT GGA TTC ACG AAG CAT GAT TCA AGC AAT GTA GAA   3120
Glu Met His Leu Arg Gly Phe Thr Lys His Asp Ser Ser Asn Val Glu
        1595                1600                1605

CAT CCG GGT ACT TTC ATT GGA GCT GTG TCG AAG CTT GAC TAT TTG AAG   3168
His Pro Gly Thr Phe Ile Gly Ala Val Ser Lys Leu Asp Tyr Leu Lys
    1610                1615                1620

GTA CAG CTG TAC TTG CTG ACT ACA TAG GAT AAT TTT TAA AGA AAG CTA   3216
Val Gln Leu Tyr Leu Leu Thr Thr  *  Asp Asn Phe  *  Arg Lys Leu
1625                1630                1635                1640

CAT ATT AGC CAG AAT TTG GGT TAT TAC AAA AAC TAC TGC ATA CTA TAG   3264
His Ile Ser Gln Asn Leu Gly Tyr Tyr Lys Asn Tyr Cys Ile Leu  *
                1645                1650                1655

CAG TTA CAT GCT CAT TAT CGA GGA GAT GCT CAC ACG CAT CTT ATT TGG   3312
Gln Leu His Ala His Tyr Arg Gly Asp Ala His Thr His Leu Ile Trp
            1660                1665                1670

ATT TAA TAC CCA ATT CTG TTT TGA TAT TGG ACT GTT CCC TCT ACA GGA   3360
Ile  *  Tyr Pro Ile Leu Phe  *  Tyr Trp Thr Val Pro Ser Thr Gly
        1675                1680                1685

GCT TGG AGT TAA TTG TAT TGA ATT AAT GCC CTG CCA TGA GTT CAA CGA   3408
Ala Trp Ser  *  Leu Tyr  *  Ile Asn Ala Leu Pro  *  Val Gln Arg
    1690                1695                1700

GCT GGA GTA CTC AAC CTC TTC TTC CAA GTA AGG ACA TGA ATT TAG TAT   3456
Ala Gly Val Leu Asn Leu Phe Phe Gln Val Arg Thr  *  Ile  *  Tyr
1705                1710                1715                1720

TAG CCT GCC AGC ACT GTT TGA GTG AGA GTT CAT ACA CAT TTT GTG CCT   3504
 *  Pro Ala Ser Thr Val  *  Val Arg Val His Thr His Phe Val Pro
                1725                1730                1735

GCA TAA CTG ATA TTT GTT CAA ACT ATT TTT TTT AGC AGT CAC TCA ACA   3552
Ala  *  Leu Ile Phe Val Gln Thr Ile Phe Phe Ser Ser His Ser Thr
            1740                1745                1750

GTT TTA CAT ATA TAT ATA ATA TAG ACT ATT CGT CAC CCT GGG TGA GGA   3600
Val Leu His Ile Tyr Ile Ile  *  Thr Ile Arg His Pro Gly  *  Gly
        1755                1760                1765

ATA GTT ATT CTT CAC CCA CCT CTA TTT TAA CAT CTA TGC ACC GTA ATT   3648
Ile Val Ile Leu His Pro Pro Leu Phe  *  His Leu Cys Thr Val Ile
    1770                1775                1780

TTA CGT TTC GTA AAT TTG TCT TAT TTT AGA GAT AAA AAG AGA ACG TAA   3696
Leu Arg Phe Val Asn Leu Ser Tyr Phe Arg Asp Lys Lys Arg Thr  *
1785                1790                1795                1800
```

GAA AAC CTA TAA TCG TCG TAA AAA AAA ATA TGT TAC GTA AAA TTA CAA 3744
Glu Asn Leu * Ser Ser * Lys Lys Ile Cys Tyr Val Lys Leu Gln
1805 1810 1815

ATG TAA AAA CAT AGT GTA AAA TGT ACA TAA AAT ACA TTT TTT GAC CTA 3792
Met * Lys His Ser Val Lys Cys Thr * Asn Thr Phe Phe Asp Leu
1820 1825 1830

TAT TTT TTT TGT TAA TGC CAA ATT TTA TAC AGT AAA TCA ATA TGA ATG 3840
Tyr Phe Phe Cys * Cys Gln Ile Leu Tyr Ser Lys Ser Ile * Met
1835 1840 1845

TAA CTA TTT GTA TTT CAA ATG TAA TTT ATT TAT GAA ATG GTC GTA AGA 3888
* Leu Phe Val Phe Gln Met * Phe Ile Tyr Glu Met Val Val Arg
1850 1855 1860

TTA CCT CGG GTG AAG AAT AAC TTA TTC TGC ACC CTG GGT GAT GAA TAG 3936
Leu Pro Arg Val Lys Asn Asn Leu Phe Cys Thr Leu Gly Asp Glu *
1865 1870 1875 1880

TAA CAC TAT ATA TAT ATA TAT ATA TAT ATA TAT ATA TAT ATA CCG GCT 3984
* His Tyr Ile Tyr Ile Tyr Ile Tyr Ile Tyr Ile Tyr Ile Pro Ala
1885 1890 1895

GCT GCT AAT GAT GTT AAT ATT TCG CAA GTA CCT AAG CTG GAT TTT TCT 4032
Ala Ala Asn Asp Val Asn Ile Ser Gln Val Pro Lys Leu Asp Phe Ser
1900 1905 1910

CCA TGA GAC ATC AAT CCA TAA TTG AAA TTG GTC ACG ACA GTT GAA TAG 4080
Pro * Asp Ile Asn Pro * Leu Lys Leu Val Thr Thr Val Glu *
1915 1920 1925

TTG ATA GCT GAA AAT GAA ATC CAG CAT GCT ACT GTC TTG CCA TCT CCA 4128
Leu Ile Ala Glu Asn Glu Ile Gln His Ala Thr Val Leu Pro Ser Pro
1930 1935 1940

GAC TTG CTA ACA TGA ATT TTG TCT GCC TAC CTG TCA TTT GTA CCA ACG 4176
Asp Leu Leu Thr * Ile Leu Ser Ala Tyr Leu Ser Phe Val Pro Thr
1945 1950 1955 1960

TTC CCA ATT GCC CTC TCA TTA TTC GTG TGT ACC ATG CAT ATG TGT TTT 4224
Phe Pro Ile Ala Leu Ser Leu Phe Val Cys Thr Met His Met Cys Phe
1965 1970 1975

AAC ATG ATT ATT GTT GGC TAT ATT TCT CTT TGG AAA CAT GAC TAA TTT 4272
Asn Met Ile Ile Val Gly Tyr Ile Ser Leu Trp Lys His Asp * Phe
1980 1985 1990

ATC ACC CGT TTT GTA TAA ACT GCT TGT TTT CAT ATC AGG ATG AAC TTT 4320
Ile Thr Arg Phe Val * Thr Ala Cys Phe His Ile Arg Met Asn Phe
1995 2000 2005

TGG GGA TAT TCT ACC ATA AAC TTC TTT TCA CCA ATG ACG AGA TAC ACA 4368
Trp Gly Tyr Ser Thr Ile Asn Phe Phe Ser Pro Met Thr Arg Tyr Thr
2010 2015 2020

TCA GGC GGG ATA AAA AAC TGT GGG CGT GAT GCC ATA AAT GAG TTC AAA 4416
Ser Gly Gly Ile Lys Asn Cys Gly Arg Asp Ala Ile Asn Glu Phe Lys
2025 2030 2035 2040

ACT TTT GTA AGA GAG GCT CAC AAA CGG GGA ATT GAG GTA AGC AAG TCG 4464
Thr Phe Val Arg Glu Ala His Lys Arg Gly Ile Glu Val Ser Lys Ser
2045 2050 2055

TAC GAG TTA GTT GCT CCT TTT GAA CTT ATC AAT TTG ATG CGA AGA CAT 4512
Tyr Glu Leu Val Ala Pro Phe Glu Leu Ile Asn Leu Met Arg Arg His
2060 2065 2070

GTT ACT GCT AGG TGA TCC TGG ATG TTG TCT TCA ACC ATA CAG CTG AGG 4560
Val Thr Ala Arg * Ser Trp Met Leu Ser Ser Thr Ile Gln Leu Arg
2075 2080 2085

GTA ATG AGA ATG GTC CAA TAT TAT CAT TTA GGG GGG TCG ATA ATA CTA 4608
Val Met Arg Met Val Gln Tyr Tyr His Leu Gly Gly Ser Ile Ile Leu
2090 2095 2100

CAT ACT ATA TGC TTG CAC CCA AGG TGA CAG ATC TTT CTT GCT GCG TAA 4656
His Thr Ile Cys Leu His Pro Arg * Gln Ile Phe Leu Ala Ala *
2105 2110 2115 2120

TTG TTC TTT CAT AGA TGT ATA GAG CAT AGA TGT GTT ATG TAG TAG TTC 4704
Leu Phe Phe His Arg Cys Ile Glu His Arg Cys Val Met * * Phe
2125 2130 2135

TTT TTC AAG GGG ATT ATG TTC ATG CAG GGA GAG TTT TAT AAC TAT TCT 4752
Phe Phe Lys Gly Ile Met Phe Met Gln Gly Glu Phe Tyr Asn Tyr Ser
2140 2145 2150

GGC TGT GGG AAT ACC TTC AAC TGT AAT CAT CCT GTG GTT CGT CAA TTC 4800
Gly Cys Gly Asn Thr Phe Asn Cys Asn His Pro Val Val Arg Gln Phe
2155 2160 2165

ATT GTA GAT TGT TTA AGG TAC AGA TAT ACA TTT TAC TTC TAG AAC TAC 4848
Ile Val Asp Cys Leu Arg Tyr Arg Tyr Thr Phe Tyr Phe * Asn Tyr
2170 2175 2180

TTT TTC ATT TCT TTT GCT GCT TGT CAT TTT GAT ATG ATT AAT TTG CAA 4896
Phe Phe Ile Ser Phe Ala Ala Cys His Phe Asp Met Ile Asn Leu Gln
2185 2190 2195 2200

GCT TGT GGG GGT AAA TCT TTT GGT CAG CAT ATT GTA TCT TTA AAT GTC 4944
Ala Cys Gly Gly Lys Ser Phe Gly Gln His Ile Val Ser Leu Asn Val
2205 2210 2215

ACA AAT ACT AAT GTC CTG GTG CTT ATT GAT TTG GCA TCT TCA AAT TCT 4992
Thr Asn Thr Asn Val Leu Val Leu Ile Asp Leu Ala Ser Ser Asn Ser
2220 2225 2230

TCT CCA ATG AAA AGG GAA AAA TCT ACT GTA TGT CTC GTC AAC TAA TTT 5040
Ser Pro Met Lys Arg Glu Lys Ser Thr Val Cys Leu Val Asn * Phe
2235 2240 2245

ACT TTT GTT TTG CAG ATA CTG GGT GAT GGA AAT GCA TGT TGA TGG TTT 5088
Thr Phe Val Leu Gln Ile Leu Gly Asp Gly Asn Ala Cys * Trp Phe
2250 2255 2260

TCG TTT TGA TCT TGC ATC CAT AAT GAC CAG AGG TTC CAG GTA ATT TGT 5136
Ser Phe * Ser Cys Ile His Asn Asp Gln Arg Phe Gln Val Ile Cys
2265 2270 2275 2280

ATT TAT TGT TTG TTT GCG TGT TGC CTT TTC AGA AGA TTC TTA AAA GAA 5184
Ile Tyr Cys Leu Phe Ala Cys Cys Leu Phe Arg Arg Phe Leu Lys Glu
2285 2290 2295

TGT TTC TTT TAC AAG TCT GTG GGA TCC AGT TAA CGT GTA TGG AGC TCC 5232
Cys Phe Phe Tyr Lys Ser Val Gly Ser Ser * Arg Val Trp Ser Ser
2300 2305 2310

```
AAT AGA AGG TGA CAT GAT CAC AAC AGG GAC ACC TCT TGT TAC TCC ACC   5280
Asn Arg Arg  *  His Asp His Asn Arg Asp Thr Ser Cys Tyr Ser Thr
        2315                2320                2325

ACT TAT TGA CAT GAT CAG CAA TGA CCC AAT TCT TGG AGG CGT CAA GGT   5328
Thr Tyr  *  His Asp Gln Gln  *  Pro Asn Ser Trp Arg Arg Gln Gly
    2330                2335                2340

ACT TGT TTC ATC CAA CAC CTG TTG TCT GTG TGC ATT CAA TTG TTT TAA   5376
Thr Cys Phe Ile Gln His Leu Leu Ser Val Cys Ile Gln Leu Phe  *
2345                2350                2355                2360

TAT GGT AAT GAT CAA TTT CCC AAT GTT GAT AAG GAA AAA AAA TGC AAG   5424
Tyr Gly Asn Asp Gln Phe Pro Asn Val Asp Lys Glu Lys Lys Cys Lys
                2365                2370                2375

TAG CTC TCT TTA TCT GCT TCT TGT GAG TTA TGC TAA ACA TGT AGA TAC   5472
 *  Leu Ser Leu Ser Ala Ser Cys Glu Leu Cys  *  Thr Cys Arg Tyr
            2380                2385                2390

TAC TAT ATT TCA ACT GTA TAT ACT TGA CAT ATT ATT GCT TCC TTG GGA   5520
Tyr Tyr Ile Ser Thr Val Tyr Thr  *  His Ile Ile Ala Ser Leu Gly
        2395                2400                2405

GGC TCT CTT ATT CCT TTC CCC CGT TGC AAT TAT AGC TCA TTG CTG AAG   5568
Gly Ser Leu Ile Pro Phe Pro Arg Cys Asn Tyr Ser Ser Leu Leu Lys
    2410                2415                2420

CAT GGG ATG CAG GAG GCC TCT ATC AAG TAG GTC AAT TCC CTC ACT GGA   5616
His Gly Met Gln Glu Ala Ser Ile Lys  *  Val Asn Ser Leu Thr Gly
2425                2430                2435                2440

ATG TTT GGT CTG AGT GGA ATG GGA AGG TAA GGT ACC TGT TAA AAG TTT   5664
Met Phe Gly Leu Ser Gly Met Gly Arg  *  Gly Thr Cys  *  Lys Phe
                2445                2450                2455

GAA TGG CAA ATA CTG ATA GAA ATA TAA CTT ATA TTT GCG ACA TAT ATA   5712
Glu Trp Gln Ile Leu Ile Glu Ile  *  Leu Ile Phe Ala Thr Tyr Ile
            2460                2465                2470

GAT AAA GCA AAA TAA TAC GCA TTC CAC CTG AAC TTT AAA GGG GCA CGC   5760
Asp Lys Ala Lys  *  Tyr Ala Phe His Leu Asn Phe Lys Gly Ala Arg
        2475                2480                2485

AGA ATT ATC CCG CAT CTG TCT ACA AGA ATG ATA ACA CAT GTG CTG AAT   5808
Arg Ile Ile Pro His Leu Ser Thr Arg Met Ile Thr His Val Leu Asn
    2490                2495                2500

AGT GAA GTA CTA CTT CTC AAA TGT CTG AAT GAA CGC ACT AAC TCT TGT   5856
Ser Glu Val Leu Leu Leu Lys Cys Leu Asn Glu Arg Thr Asn Ser Cys
2505                2510                2515                2520

GAG TGT CAA CCG AGC AAG AAA TAT TTG AGT TTT CTG CAA GAA ATT GTT   5904
Glu Cys Gln Pro Ser Lys Lys Tyr Leu Ser Phe Leu Gln Glu Ile Val
                2525                2530                2535

CAT GTT GTG CTG TAT TAT ACT CCC TCC GTC CGA AAT TAT TTG TCG GAG   5952
His Val Val Leu Tyr Tyr Thr Pro Ser Val Arg Asn Tyr Leu Ser Glu
            2540                2545                2550

AAA TGG ATG TAT CTA GAC GTA TTT TAG TTC TAG ATA CAT CCA TTT TTA   6000
Lys Trp Met Tyr Leu Asp Val Phe  *  Phe  *  Ile His Pro Phe Leu
        2555                2560                2565
```

```
TCC ATT TCT GCA ACA AGT AGT TCC GGA CGG AGG GAG TAT CAT TTA ACA    6048
Ser Ile Ser Ala Thr Ser Ser Ser Gly Arg Arg Glu Tyr His Leu Thr
    2570                2575                2580

AAT ATA TGC ATG TTC GAA GTA AAT CCC CAC GAA TAA GCA TAT AAG ACG    6096
Asn Ile Cys Met Phe Glu Val Asn Pro His Glu  *  Ala Tyr Lys Thr
2585                2590                2595                2600

ATA TTG CTT TTT GAC TTG CAA CAC CTA AAC CTC ATT GTT TTC TCC TAG    6144
Ile Leu Leu Phe Asp Leu Gln His Leu Asn Leu Ile Val Phe Ser  *
                2605                2610                2615

GAT TTT GGG TGT TCG AAG CAA GCA GCT GGT GAT ATT TAA TTT ACC TTT    6192
Asp Phe Gly Cys Ser Lys Gln Ala Ala Gly Asp Ile  *  Phe Thr Phe
            2620                2625                2630

GCC TTT ATT TGT AGC TTG ATT TGA GGG TGC GGC AAA GGT TTT AGC TTA    6240
Ala Phe Ile Cys Ser Leu Ile  *  Gly Cys Gly Lys Gly Phe Ser Leu
        2635                2640                2645

GTA GTG TTT TGT AAA TTA TTA TAG TTT ATG TAT ATA CTC CTC ATT TGG    6288
Val Val Phe Cys Lys Leu Leu  *  Phe Met Tyr Ile Leu Leu Ile Trp
    2650                2655                2660

GCA CTT CCG TAC TGG TCC CAT AGA AGA TAA AAA TGG AAT GAT GTC TGG    6336
Ala Leu Pro Tyr Trp Ser His Arg Arg  *  Lys Trp Asn Asp Val Trp
2665                2670                2675                2680

CCA ATA ATT GTT GAC AAC ACT GTT GCG CAT TTG ATT TTT ATC AGG GAA    6384
Pro Ile Ile Val Asp Asn Thr Val Ala His Leu Ile Phe Ile Arg Glu
                2685                2690                2695

TGG AAA ATT GAA ATC GGT AAG AAA CAT TGC GAT ATT AAG CTT GTA TAT    6432
Trp Lys Ile Glu Ile Gly Lys Lys His Cys Asp Ile Lys Leu Val Tyr
            2700                2705                2710

GCT AAT GCT GGT GGA TCT TTA AGA GGG AAC ATA TGA TCT CGT GTG CAT    6480
Ala Asn Ala Gly Gly Ser Leu Arg Gly Asn Ile  *  Ser Arg Val His
        2715                2720                2725

CCA TCT TCA ACT AAA AAA ATA TGT TGC ACA TCT CCC ACG TCA CTT ACT    6528
Pro Ser Ser Thr Lys Lys Ile Cys Cys Thr Ser Pro Thr Ser Leu Thr
    2730                2735                2740

AGC TAT TTC ATC CAA GTA CTA ACT TGT GTG GTT GTC TCC TCA GTA CCG    6576
Ser Tyr Phe Ile Gln Val Leu Thr Cys Val Val Val Ser Ser Val Pro
2745                2750                2755                2760

GGA CAT TGT GCG CCA ATT CAT TAA AGG CAC TGA TGG ATT TGC TGG TGG    6624
Gly His Cys Ala Pro Ile His  *  Arg His  *  Trp Ile Cys Trp Trp
                2765                2770                2775

TTT TGC CGA ATG TCT TTG TGG AAG TCC ACA CCT ATA CCA GGT AAG TTG    6672
Phe Cys Arg Met Ser Leu Trp Lys Ser Thr Pro Ile Pro Gly Lys Leu
            2780                2785                2790

TGG CAA TAC TTG GAA ATG GGT TGA GTG AAT GTC ACA TGG ATT TTT TAT    6720
Trp Gln Tyr Leu Glu Met Gly  *  Val Asn Val Thr Trp Ile Phe Tyr
        2795                2800                2805

ATA TAC CAC ATG ATG ATA CAC ATG TAA ATA TAT AAC GAT TAT AGT GTA    6768
Ile Tyr His Met Met Ile His Met  *  Ile Tyr Asn Asp Tyr Ser Val
    2810                2815                2820
```

```
TGC ATA TGC ATT TGG CTA AGA AGT ACT CCC TCC CTT AGT AAA AGT TAG   6816
Cys Ile Cys Ile Trp Leu Arg Ser Thr Pro Ser Leu Ser Lys Ser  *
2825                2830                2835                2840

TAC AAA GTT GAG TCA TCT ATT TTG GAA CGG AGG GAG TAT AAG TGT ATA   6864
Tyr Lys Val Glu Ser Ser Ile Leu Glu Arg Arg Glu Tyr Lys Cys Ile
            2845                2850                2855

CAC TAG TGC AAT ATA TAG GTT TTA ACA CCC AAC TTG CCA ATG AAG GAA   6912
His  *  Cys Asn Ile  *  Val Leu Thr Pro Asn Leu Pro Met Lys Glu
        2860                2865                2870

CAT AGG GCT TTC TAG TTA TCT TAT TTA TTT GTC TGG TGA ATA ATC CAC   6960
His Arg Ala Phe  *  Leu Ser Tyr Leu Phe Val Trp  *  Ile Ile His
    2875                2880                2885

TGA AAA ATT CCA GCC ATG TCA TTT TTT AGG GGG GGA GAA GAA ACT ACA   7008
 *  Lys Ile Pro Ala Met Ser Phe Phe Arg Gly Gly Glu Glu Thr Thr
    2890                2895                2900

TTG ATT TTT CCC CCT AAA AAA AGC CAT CTC AGA TTT CAT AGG TAA CTT   7056
Leu Ile Phe Pro Pro Lys Lys Ser His Leu Arg Phe His Arg  *  Leu
2905                2910                2915                2920

GCT TTT CTG TAA AGA AAT GAA AAC GAC TTC ATA CTT TCT GTC GAT TAT   7104
Ala Phe Leu  *  Arg Asn Glu Asn Asp Phe Ile Leu Ser Val Asp Tyr
            2925                2930                2935

AAG TGT ATA CAC TAG TGC AAT ATA TAG GTT TTA ACA CCC AAC TTG CCA   7152
Lys Cys Ile His  *  Cys Asn Ile  *  Val Leu Thr Pro Asn Leu Pro
        2940                2945                2950

ATG AAG GAA CAT AGG GCT TTC TAG TTA TCT TAT TTA TTT GCT GGT GAA   7200
Met Lys Glu His Arg Ala Phe  *  Leu Ser Tyr Leu Phe Ala Gly Glu
    2955                2960                2965

TAA TCC ACT GAA AAA TTC CAG CCA TGT CAT TTT TTA GGG GGG AGA AGA   7248
 *  Ser Thr Glu Lys Phe Gln Pro Cys His Phe Leu Gly Gly Arg Arg
    2970                2975                2980

AAC TAT ATT GAT TTT TCC CCC TAA AAA AAG CCA TCT CAG ATT CAT AGG   7296
Asn Tyr Ile Asp Phe Ser Pro  *  Lys Lys Pro Ser Gln Ile His Arg
2985                2990                2995                3000

AAC TTG CTT TTC TGT AAA GAA ATG AAA ACG ACT TCA TAC TTT CTG CGG   7344
Asn Leu Leu Phe Cys Lys Glu Met Lys Thr Thr Ser Tyr Phe Leu Arg
            3005                3010                3015

CGC TTA CTT AGC TCG ATG GAT ATT TGT AAG ATG AAT GCC AAA TTA TTT   7392
Arg Leu Leu Ser Ser Met Asp Ile Cys Lys Met Asn Ala Lys Leu Phe
        3020                3025                3030

GGC GGG ATT TGA TCG TTA TTC CAA ATT TCA TTT GGT TTC TCT AGC AAT   7440
Gly Gly Ile  *  Ser Leu Phe Gln Ile Ser Phe Gly Phe Ser Ser Asn
    3035                3040                3045

CAA CCC AGT ACC TTG TTA TTG GCA CTG CAA TTT CTT ATT GAT TAA TCA   7488
Gln Pro Ser Thr Leu Leu Leu Ala Leu Gln Phe Leu Ile Asp  *  Ser
    3050                3055                3060

GGC AGG AGG AAG GAA ACC TTG GCA CAG TAT CAA CTT GGT ATG TGC ACA   7536
Gly Arg Arg Lys Glu Thr Leu Ala Gln Tyr Gln Leu Gly Met Cys Thr
3065                3070                3075                3080
```

```
TGA TGG ATT TAC ACT GGG TGA TTT GGT ACA TAT AAT ACC AAG TCA ATT    7584
 *  Trp Ile Tyr Thr Gly  *  Phe Gly Thr Tyr Asn Thr Lys Ser Ile
                3085                3090                3095

TAC CAA ATG GGG AGA CCA ATA GAG ATG GAG AAA ATC ACA ATC TTA GCT    7632
Tyr Gln Met Gly Arg Pro Ile Glu Met Glu Lys Ile Thr Ile Leu Ala
            3100                3105                3110

GGA ATT GTG GGG AGG TAA TTC TGA ACT CTC CTT TTT TTT TGA AAT TTT    7680
Gly Ile Val Gly Arg  *  Phe  *  Thr Leu Leu Phe Phe  *  Asn Phe
        3115                3120                3125

CAT GCT TTA CAT AAT AGT CAA ATG GCT GAC AAA TGT CGT TGT ATG GTT    7728
His Ala Leu His Asn Ser Gln Met Ala Asp Lys Cys Arg Cys Met Val
    3130                3135                3140

CTC TCT ACC TAA ACC GTT AAG GCA GTA AGA GTT TCC CTA CAA GAT CTC    7776
Leu Ser Thr  *  Thr Val Lys Ala Val Arg Val Ser Leu Gln Asp Leu
3145                3150                3155                3160

TTT GTT CGT ATA ATT GTA TTT TCT AGA GAA AAG TTG CCT TCA ATT TTG    7824
Phe Val Arg Ile Ile Val Phe Ser Arg Glu Lys Leu Pro Ser Ile Leu
                3165                3170                3175

TGC ACG CGG CAG TAC AGG AAT TGT GGT TAT AAA TAT TGA TAC AGG CTG    7872
Cys Thr Arg Gln Tyr Arg Asn Cys Gly Tyr Lys Tyr  *  Tyr Arg Leu
            3180                3185                3190

ACC ATC GTT ACT AAT AGG GGG AAC AAT AAG CAC ATT TTT TTA ATA GCA    7920
Thr Ile Val Thr Asn Arg Gly Asn Asn Lys His Ile Phe Leu Ile Ala
        3195                3200                3205

AAG GCA TCA CCC TTG TTC CGT TTC CAA TGA AAT CAC AGT ATC CGA ACC    7968
Lys Ala Ser Pro Leu Phe Arg Phe Gln  *  Asn His Ser Ile Arg Thr
    3210                3215                3220

ATA AGT TTT ACA AGT ATG CGT AGA GAG AAA TAA AGT ATC AAC CCG GCA    8016
Ile Ser Phe Thr Ser Met Arg Arg Glu Lys  *  Ser Ile Asn Pro Ala
3225                3230                3235                3240

GAA ACA GTT GTT TCA GGC GCA AAG AGA AAA GGA AAC GAT ATG CTC TAT    8064
Glu Thr Val Val Ser Gly Ala Lys Arg Lys Gly Asn Asp Met Leu Tyr
                3245                3250                3255

TAC ATC AAC CTT TTA GCA TTT AGG GAC GAC CAG CAT CAT CCC ATC TTC    8112
Tyr Ile Asn Leu Leu Ala Phe Arg Asp Asp Gln His His Pro Ile Phe
            3260                3265                3270

AAT CAA CTG GAG CGA GGT CAC CTC CAA TCT TCT CAG CAG CCT CAG AGT    8160
Asn Gln Leu Glu Arg Gly His Leu Gln Ser Ser Gln Gln Pro Gln Ser
        3275                3280                3285

GGT GAC CTC CCA AGC AAG TGC ATC AGC ATC CAT CAT CTG GGG GTT GGG    8208
Gly Asp Leu Pro Ser Lys Cys Ile Ser Ile His His Leu Gly Val Gly
    3290                3295                3300

CAC ATA CCA TGA GCA CAA TCA CCT GAA TTT GAT GAA TTT TCC TCT GTT    8256
His Ile Pro  *  Ala Gln Ser Pro Glu Phe Asp Glu Phe Ser Ser Val
3305                3310                3315                3320

TAC CTT GCA GCA GAC CCC TGC CGT ATA AAT GGT TTT AAA TGA CAG CAT    8304
Tyr Leu Ala Ala Asp Pro Cys Arg Ile Asn Gly Phe Lys  *  Gln His
                3325                3330                3335
```

GTT CTT TCA GTT TGA GCA AAA TTT GTG CAA TTG CAA AGA AGC TTT AGA 8352
Val Leu Ser Val * Ala Lys Phe Val Gln Leu Gln Arg Ser Phe Arg
3340 3345 3350

ATC ATG TGG AAC ATG CAC TTA CAT TTC ATC TGA CAA TAT AGG AAG GAG 8400
Ile Met Trp Asn Met His Leu His Phe Ile * Gln Tyr Arg Lys Glu
3355 3360 3365

AGC CCG ACG TCG CAT GCT CCT CTA GAC TCG AGG AAT TCG CAA GAT TGT 8448
Ser Pro Thr Ser His Ala Pro Leu Asp Ser Arg Asn Ser Gln Asp Cys
3370 3375 3380

CTG TCA AAA GAT TGA GGA AGA GGC AGA TGC GCA ATT TCT TTG TTT GTC 8496
Leu Ser Lys Asp * Gly Arg Gly Arg Cys Ala Ile Ser Leu Phe Val
3385 3390 3395 3400

TCA TGG TTT CTC AAG TAA GAC TTA TAT CTG ATC TCT TCA ATT TTT GAG 8544
Ser Trp Phe Leu Lys * Asp Leu Tyr Leu Ile Ser Ser Ile Phe Glu
3405 3410 3415

ATT GCC TGT TTT TCA CAA TGG CAT ATG TTG TCA GGT GAA ACA TCC AAT 8592
Ile Ala Cys Phe Ser Gln Trp His Met Leu Ser Gly Glu Thr Ser Asn
3420 3425 3430

CCC AGT ATT AAT AGA GCC AAC ATG AAG GGA TTG CTT ATC TGA GAT ATC 8640
Pro Ser Ile Asn Arg Ala Asn Met Lys Gly Leu Leu Ile * Asp Ile
3435 3440 3445

TGC CAA AGT TGA ATT CTT AGA TTC ACC TTC TTC AGT ATT TCA GAC CTT 8688
Cys Gln Ser * Ile Leu Arg Phe Thr Phe Phe Ser Ile Ser Asp Leu
3450 3455 3460

CTA AGC ATT TTC ATT TTT TTT TTC AAT TGT TAG GGA GTT CCA ATG TTT 8736
Leu Ser Ile Phe Ile Phe Phe Phe Asn Cys * Gly Val Pro Met Phe
3465 3470 3475 3480

TAC ATG GGC GAT GAA TAT GGC CAC ACA AAA GGG GGC AAC AAC AAT ACA 8784
Tyr Met Gly Asp Glu Tyr Gly His Thr Lys Gly Gly Asn Asn Asn Thr
3485 3490 3495

TAC TGC CAT GAT TCT TAT GTC AGT ACA ATT TGG TCA CAT ATT GTT GTT 8832
Tyr Cys His Asp Ser Tyr Val Ser Thr Ile Trp Ser His Ile Val Val
3500 3505 3510

CTA AGT AAC TAT CTT CAA ATC TTT GCA TTC ATC CGT CAT GGC TCT TCT 8880
Leu Ser Asn Tyr Leu Gln Ile Phe Ala Phe Ile Arg His Gly Ser Ser
3515 3520 3525

GTA GGT CAA TTA TTT TCG CTG GGA TAA AAA AGA ACA ATA CTC TGA CTT 8928
Val Gly Gln Leu Phe Ser Leu Gly * Lys Arg Thr Ile Leu * Leu
3530 3535 3540

GCA AAG ATT CTG CTG CCT CAT GAC CAA ATT CCG CAA GTA AGT ATT CCG 8976
Ala Lys Ile Leu Leu Pro His Asp Gln Ile Pro Gln Val Ser Ile Pro
3545 3550 3555 3560

TTG AAT AAT TTC TGT GTA GAA CCA CTG AAG GTG CCT CCA AAC GCT AAG 9024
Leu Asn Asn Phe Cys Val Glu Pro Leu Lys Val Pro Pro Asn Ala Lys
3565 3570 3575

CGA GCA AGG TCA ATT TCA CAC CCT AAT CAA GTT GGT GTT GTC TAT TTG 9072
Arg Ala Arg Ser Ile Ser His Pro Asn Gln Val Gly Val Val Tyr Leu
3580 3585 3590

```
TGT ATT TGA TCT GCT GCA CTG TAG GGA GTG CGA GGG TCT TGG CCT TGA    9120
Cys Ile  *  Ser Ala Ala Leu  *  Gly Val Arg Gly Ser Trp Pro  *
        3595                3600                3605

GGA CTT TCC AAC GGC CGA ACG GCT GCA GTG GCA TGG TCA TCA GCC TGG    9168
Gly Leu Ser Asn Gly Arg Thr Ala Ala Val Ala Trp Ser Ser Ala Trp
    3610                3615                3620

GAA GCC TGA TTG GTC TGA GAA TAG CCG ATT CGT TGC CTT TTC CAT GGT    9216
Glu Ala  *  Leu Val  *  Glu  *  Pro Ile Arg Cys Leu Phe His Gly
3625                3630                3635                3640

ACA CAT ATA GTT CTG ACA CTT CAC TAT AGT TGT TTT AAA AAA GAA AAT    9264
Thr His Ile Val Leu Thr Leu His Tyr Ser Cys Phe Lys Lys Glu Asn
                3645                3650                3655

TTA ACT CAA AAG TAA ATT ATG GAG  A                                 9289
Leu Thr Gln Lys  *  Ile Met Glu
            3660
```

CLAIMS

1. A nucleic acid sequence encoding an enzyme of the starch biosynthetic pathway in a cereal plant, wherein the enzyme is selected from the group consisting of starch branching enzyme I, starch branching enzyme II, starch soluble synthase I, and debranching enzyme, with the proviso that the enzyme is not soluble starch synthase I of rice, or starch branching enzyme I of rice or maize.

2. A sequence according to claim 1, wherein the sequence is a genomic DNA or cDNA sequence.

3. A sequence according to claim 1 or claim 2, wherein the sequence is functional in wheat.

4. A sequence according to any one of claims 1 to 3, wherein the sequence is derived from a *Triticum* species.

5. A sequence according to claim 4, wherein the *Triticum* species is *Triticum tauschii*.

6. A sequence according to any one of claims 1 to 5, wherein the sequence encodes starch branching enzyme I or a biologically-active fragment thereof, and wherein the sequence has at least 70% sequence homology with the sequence shown in SEQ ID NO:5 or SEQ ID NO:9.

7. A sequence according to claim 6, wherein the homology is at least 90%.

8. A sequence according to any one of claims 1 to 5, wherein the sequence encodes starch branching enzyme II a or biologically-active fragment thereof, and wherein the sequence has at least 70% sequence homology with the sequence shown in SEQ ID NO:10.

9. A sequence according to claim 8, wherein the homology is at least 90%.

10. A sequence according to any one of claims 1 to 5, wherein the sequence encodes soluble starch synthase or a biologically-active fragment thereof, and wherein the sequence has at least 70% sequence homology with the sequence shown in SEQ ID NO:11 or SEQ ID NO:13.

11. A sequence according to claim 10, wherein the homology is at least 90%.

12. A sequence according to claim 11, wherein the sequence encodes a 75 kD soluble starch synthase of wheat.

13. A sequence according to claim 12, which encodes an amino acid sequence at least 70% homologous to that shown in SEQ ID NO:14.

14. A sequence according to any one of claims 1 to 5, wherein the sequence encodes debranching enzyme or a biologically-active fragment thereof, and wherein the sequence has at least 70% sequence homology with the sequence shown in SEQ ID No:17.

15. A sequence according to claim 14, wherein the homology is at least 90%.

16. A promoter of an enzyme selected from the group consisting of starch branching enzyme I, starch branching enzyme II, starch soluble synthase I, and debranching enzyme, with the proviso that the enzyme is not soluble starch synthase I of rice, or starch branching enzyme I of rice or maize.

17. A promoter according to claim 16, wherein the promoter is a starch branching enzyme I promoter or

biologically-active fragment thereof, and wherein the promoter sequence has at least 70% sequence homology with the sequence shown in SEQ ID No:8.

18. A sequence according to claim 17, wherein the homology is at least 90%.

19. A promoter according to claim 16, wherein the promoter is a starch soluble synthase I promoter or biologically-active fragment thereof, and wherein the promoter sequence has at least 70% sequence homology with the sequence shown in SEQ ID No:15.

20. A sequence according to claim 19, wherein the homology is at least 90%.

21. A nucleic acid construct comprising a nucleic acid sequence encoding an enzyme of the starch biosynthetic pathway in a cereal plant, operably linked to one or more nucleic acid sequences facilitating expression of the nucleic acid sequence in a plant, wherein the enzyme is selected from the group consisting of starch branching enzyme I, starch branching enzyme II, starch soluble synthase I, and debranching enzyme, with the proviso that the enzyme is not soluble starch synthase I of rice, or starch branching enzyme I of rice or maize, a biologically-active fragment thereof.

22. A nucleic acid construct for targeting a gene to the endosperm of a cereal plant, comprising one or more promoter sequences selected from the group consisting of SBE I promoter, SBE II promoter, SSS I promoter, and DBE promoter, operatively linked to a nucleic acid sequence encoding a protein, wherein the expression of the targetted gene in the endosperm of a cereal plant is modified.

23. A construct according to either claim 21 or claim 22, wherein the promoter or nucleic acid sequence is also operatively linked to one or more additional targeting sequences and/or one or more 3' untranslated sequences.

24. A construct according to claim 23, wherein the nucleic acid encoding the protein is either in the sense or antisense orientation.

25. A construct according to claims 24, wherein the protein is an enzyme of the starch biosynthetic pathway.

26. A construct according to claim 25, wherein the nucleic acid encoding the protein is in the antisense orientation, and the enzyme is selected from the group consisting of GBSS, starch debranching enzyme, SBE II, low molecular weight glutenin, and grain softness protein I.

27. A construct according to claim 25, wherein the nucleic acid encoding the protein is in the sense orientation, and the enzyme is selected from the group consisting of bacterial isoamylase, bacterial glycogen synthase, and wheat high molecular weight glutenin Bx17.

28. A construct according to any one of claims 21 to 27, wherein the plant is a cereal plant.

29. A construct according to claim 28, wherein the cereal plant is either wheat or barley.

30. A construct according to claim 29, wherein the cereal plant is wheat.

31. A construct according to any one of claims 21 to 30, wherein the construct is either a plasmid or a vector.

32. A construct according to claim 31, wherein the plasmid or vector is suitable for use in the transformation of a plant.

33. A construct according to claim 32, wherein the plasmid is selected from the group consisting of those depicted in Figures 22a to 22f.

34. A construct according to claim 32, wherein the vector is a bacterium of the genus *Agrobacterium*.

35. A construct according to claim 34, wherein the vector is *Agrobacterium tumefaciens*.

36. A method of modifying the characteristics of starch produced by a plant, comprising the steps of:

(a) introducing a nucleic acid sequence encoding an enzyme of the starch biosynthetic pathway into a host plant, and/or

(b) introducing an anti-sense nucleic acid sequence directed to a gene encoding an enzyme of the starch biosynthetic pathway into a host plant,

wherein the enzyme is selected from the group consisting of starch branching enzyme I, starch branching enzyme II, starch soluble synthase I, and debranching enzyme, with the proviso that the enzyme is not soluble starch synthase I of rice, or starch branching enzyme I of rice or maize, and wherein if both steps (a) and (b) are used, the enzymes in the two steps are different.

37. A method according to claim 36, wherein the plant is a cereal plant.

38. A method according to claim 37, wherein the cereal plant is wheat or barley.

39. A method of targeting expression of a gene to the endosperm of a cereal plant, comprising the step of transforming the plant with a construct according to any one of claims 21 to 35.

40. A method of modulating the time of expression of a gene in endosperm of a cereal plant, comprising the step of transforming the plant with a construct according to any one of claims 21 to 35.

41. A method according to claim 40, wherein when expression at an early stage following anthesis is desired, the construct comprises either the SBE II, SSS I, or DBE promoter.

42. A method according to claim 40, wherein when expression at a later stage following anthesis is desired, the construct comprises the SBE I promoter.

43. A plant transformed with a construct according to any one of claims 21 to 35.

44. A plant according to claim 43, wherein the plant is a cereal plant.

45. A plant according to claim 44, wherein the cereal plant is wheat or barley.

46. A method of identifying variations in the starch synthesis characteristics of a cereal plant, comprising the step of identifying a variation in nucleic acid sequence in the intron regions of the SBE I, SBE II, SSS I or DBE genes.

47. A method of identifying variations in the starch synthesis characteristics of a cereal plant, comprising the step of identifying a variation in nucleic acid sequence compared to the sequence shown in one or more SEQ ID NO:5,

SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, or SEQ ID NO:17.

48. A method according to claim 47, in which a mutation or absence of a SBE I, SBE II, SSS I or DBE gene is detected.

49. A method according to either claim 47 or claim 48, in which the cereal plant is wheat or barley.

50. A product comprising plant material propogated from a plant transformed with a nucleic acid sequence encoding an enzyme of the starch biosynthetic pathway in a cereal plant, operably linked to one or more nucleic acid sequences facilitating expression of the nucleic acid sequence in a plant, wherein the enzyme is selected from the group consisting of starch branching enzyme I, starch branching enzyme II, starch soluble synthase I, and debranching enzyme, with the proviso that the enzyme is not soluble starch synthase I of rice, or starch branching enzyme I of rice or maize, a biologically-active fragment thereof.

51. A product comprising plant material propogated from a plant in which a gene was targeted to the endosperm of a cereal plant, by a nucleic acid construct comprising one or more promoter sequences selected from the group consisting of SBE I promoter, SBE II promoter, SSS I promoter, and DBE promoter, operatively linked to a nucleic acid sequence encoding a protein, wherein the expression of the targetted gene in the endosperm of a cereal plant is modified.

52. A product according to claim 50 or claim 51 wherein the product is a food product.

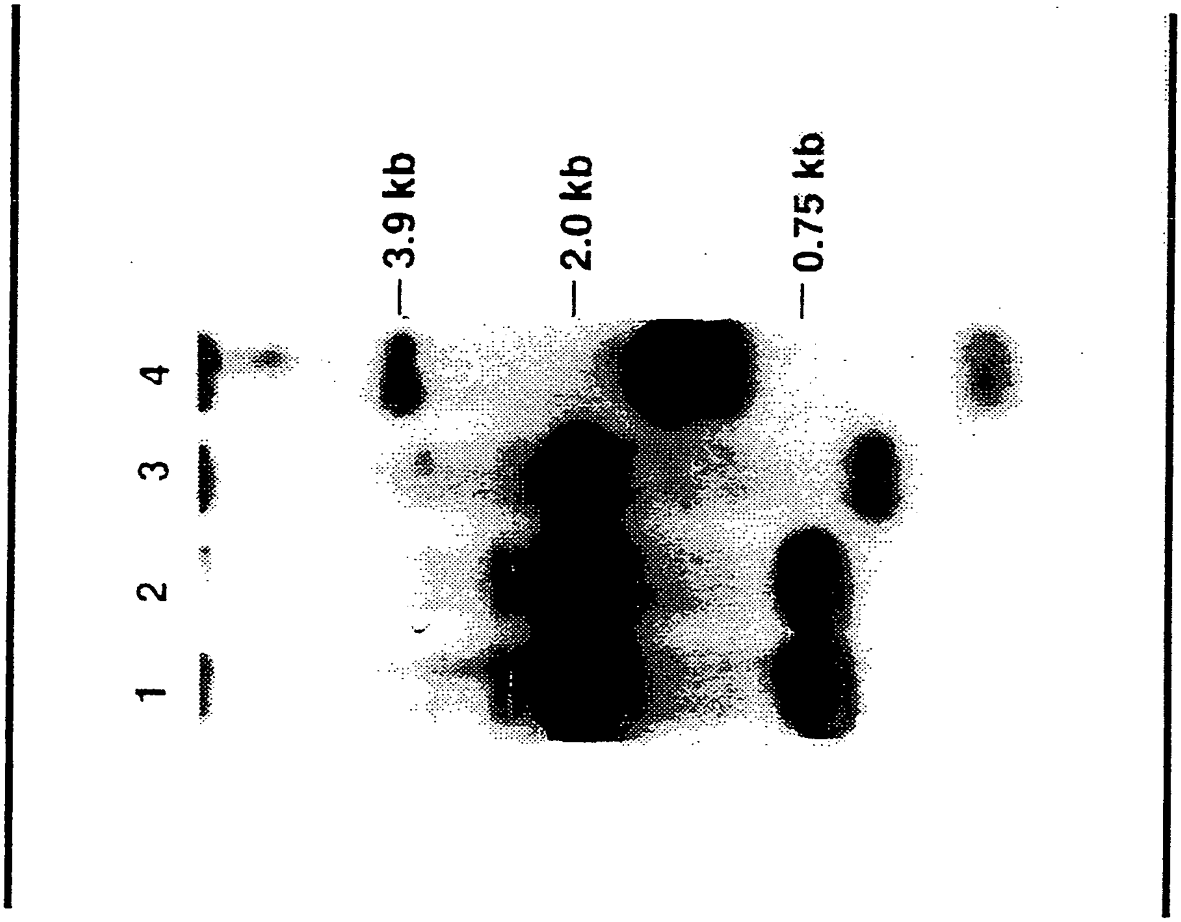


FIGURE 1

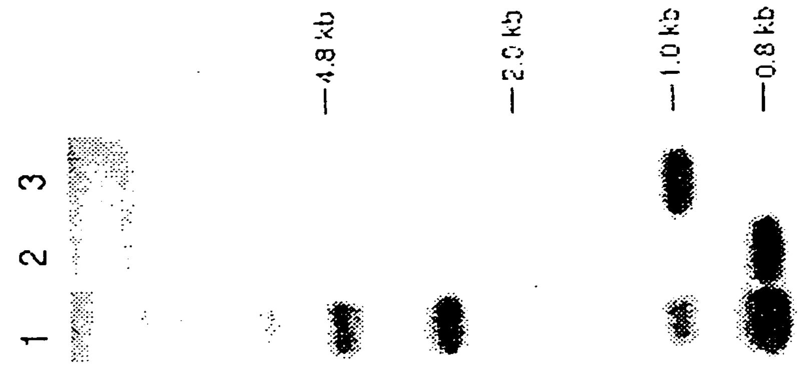


FIGURE 2

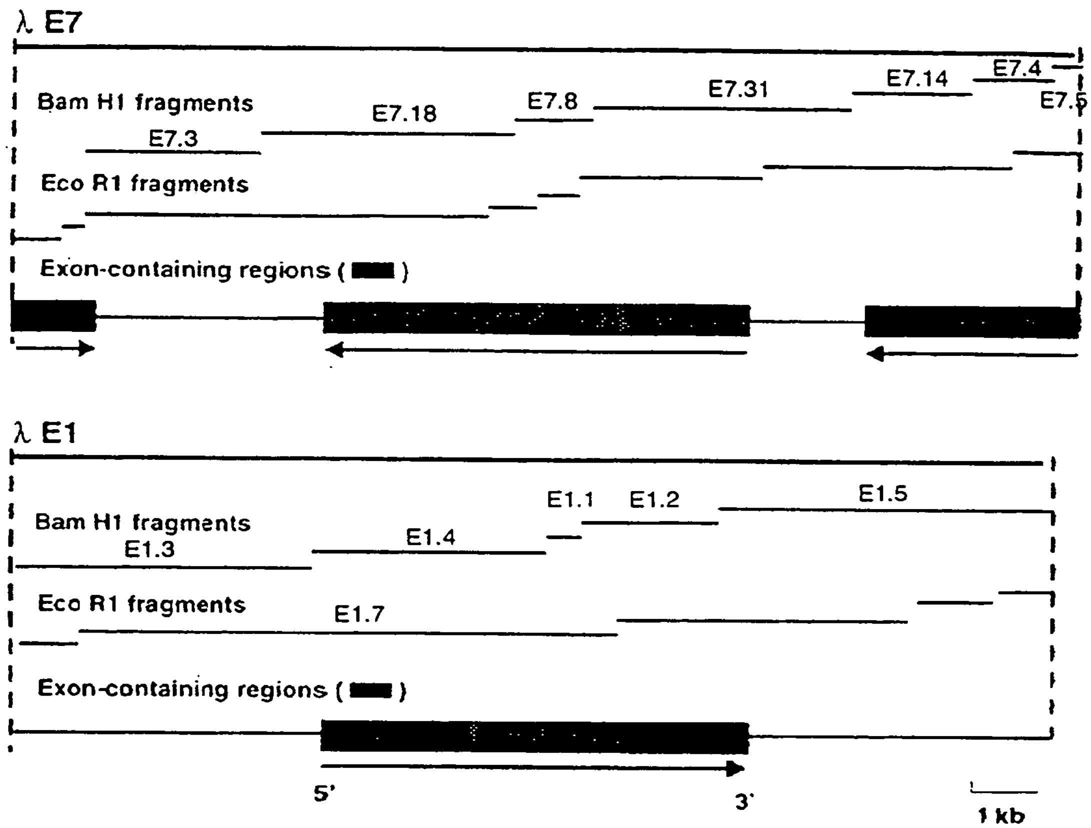

λ E7
Bam H1 fragments
E7.3
E7.18
E7.8
E7.31
E7.14
E7.4
E7.5
Eco R1 fragments
Exon-containing regions ( )
λ E1
Bam H1 fragments
E1.3
E1.4
E1.1
E1.2
E1.5
Eco R1 fragments
E1.7
Exon-containing regions ( )
5'
3'
1 kb

FIGURE 3

4/44

```
           1                                                      50
RSBEI      .......... .********* *....***pl lp******** **ag******
MSBEI      .......... .****v*p** **tplp***r ***h***aa* pg********
D4cDNA     .......... .*****ap*c **sl..***p **pa****g* **s*......
PESBEII    .......... .......... .......... .......... ..........
POSBE      meinfkvlsk pirgsfp*f* pkv*sgas*n kic*psqh*t *lkf*sqers
D2cDNA     .......... .****s**ll prp*a*.... ....****l* *******ggk
Consensus  ---------- -MLCLTSSSS SP-S-APPR- SRS-ADRPSP GIIAGGGNVR

           51                                                    100
RSBEI      l..**v*... *p*****g** *tn***pa** rk****v*vv ***..*****
MSBEI      l..**l**qc ka***gv*** ****ataa*v q*d*****ak g**..*****
D4cDNA     .......... .......... ******p*s* prdy****a* *g*..gd***
PESBEII    .......... .......... ........mt d*ks**psv* **f..*nig*
POSBE      w..d*s*t*k *rv*kde*mk h*saisa*lt d**s***pl* ***kt*nigl
D2cDNA     rlsv*p***f ll**l****a ***sf*s*** rg**ia**.. tgygs*****
Consensus  ---SV-SVP- S-RRSWPRKV KSKFSV-VTA -DNKTMAT-E EDV--DHLPI

           101                                                   150
RSBEI      ********e* ****n**i** *****c**** ********** *********v
MSBEI      ********i* ********** *****gs**e n**s**s*** *********n
D4cDNA     *******ag* ****s****k ******s*** ********** **********
PESBEII    lnv*ss**p* ****k***** **h**k***e y****q**a* ******f*r*
POSBE      ln***t**p* l****h**** *v***m**** y**p****aq ******f*r*
D2cDNA     ****l**ae* ****d*trn* *i******** ***s****** **********
Consensus  YDLDPKLE-F KDHFRYRMKR YLDQKHLIEK HEGGLEEFSK GYLKFGINTE

           151                                                   200
RSBEI      *g******** ********** *******ak* *****k**** **k*******
MSBEI      *dg******* ******e*** ***d***a** *****k**** **k*d**k**
D4cDNA     nd******** ***m****** ********g* r*t**n**** **********
PESBEII    *dgis***** ********i** ***g*****l h****q**** **q*pdad*n
POSBE      *gci****** *****dev** ***g****** m****q**** ****pd*ds*
D2cDNA     hg*s****** ***e****** ********g* **a**n**** **********
Consensus  --ATVYREWA PAAQEAQLIG DFNNWNGSNH KMEKD-FGVW SIRISHVNGK

           201                                                   250
RSBEI      ********** ***r**g*a* ********** **f******* **********
MSBEI      ********** ***l*.g*** ******l*** ********** **********
D4cDNA     ********** ***hr*d*l* ********** **f******* **********
PESBEII    *******r** ***k*sd*** ********k* ****ptr*a* *****y****
POSBE      *v*****r** ***k**n*** ********k* **a**t**a* *****y****
D2cDNA     ********** ***r*.h*** **q******* ***t**es** ****l*****
Consensus  PAIPHNSKVK FRF-HG-GVW VDRIPAWIRY ATVDASKFGA PYDGVHWDPP

           251                                                   300
RSBEI      ac******** ********** ********** ********** **********
MSBEI      a****t**** **s**a**** ********** k*a******* **********
D4cDNA     sg******** **r******* ********** r********* ********k*
PESBEII    l****q**** *****k**** ********ss **r*ns**** **d******e
POSBE      p****h**y* *****r**** ********ss **r*ns**** **d*****k*
D2cDNA     s******n** ******v*** ******v**g kl*ag***** p*****cl**
Consensus  -SERYVFKHP RPPKPDAPRI YEAHVGMSGE EPEVSTYREF ADNVLPRIRA
```

Figure 4

```
           301                                                  350
RSBEI      ********** ********** ******.*** ********** **********
MSBEI      ********** ********** ******.*** ********** **********
D4cDNA     ********** *****ilcf* w*****.*** ********** **********
PESBEII    ********** ********** w****kp*** ******s*** **********
POSBE      ********** *******g** ******.*** ****y*n*** **********
D2cDNA     t********g *****ds*** ******.*** ********** **********
Consensus  NNYNTVQLMA IMEHSYYASF GYHVTN-FFA VSSRSGTPED LKYL-DKAHS

           351                                                  400
RSBEI      ********** ********** *********n *h*****t** **********
MSBEI      ********** ********** ********** *******a** **********
D4cDNA     ********** *****s*m** *********n *******t** **********
PESBEII    ***n****** ********** ********** s*q****a** **********
POSBE      ***q**v*** ********** *********g s******a** **********
D2cDNA     ********** ********i* ********** ah****yt** k**n***ng*
Consensus  LGLRVLMDVV HSHASNNVTD GLNGYDVGQS TQESYFH-GD RGYHKLWDSR

           401                                                  450
RSBEI      ********** ********** ********** ********** *****k****
MSBEI      ********** ********** ********** ********** *****v****
D4cDNA     ********** ********** ********** *********n ******s*a*
PESBEII    *******ks. s********* ****k***** ********** ******a***
POSBE      ********** ********** ****n***** *********v **********
D2cDNA     ********** ********** ********** *v*******n *n****s*n*
Consensus  LFNYANWEVL RFLLSNLRYW -DEFMFDGFR FDGVTSMLYH HHGINMGFTG

           451                                                  500
RSBEI      ********** ********** *******l** ********** **********
MSBEI      **q******* a********* *******l** ********** **********
D4cDNA     ******g*** ********** *******i** ********** *******s**
PESBEII    d*n****e** ********** **s*v*di** ***d****** ***g*g***s
POSBE      **n****ea* ********** **n*i**i** ********** ***g*g***s
D2cDNA     *****ig*** n***f***** *******l** **i***v*** **********
Consensus  NYKEYFSLDT DVDAVVYMML ANHLMHK-LP EATVVAEDVS GMPVLCRPVD

           501                                                  550
RSBEI      ********** ********** *******rk* ****.*vq** **********
MSBEI      ********** ********** ********** **g*.*ah** **********
D4cDNA     ********** ********** *******l** ***a.*ah** **********
PESBEII    *v******** ******k*** *****k**** **k*.*sln* **********
POSBE      ********** ******k*** *****n*e** **k*.*tss* **********
D2cDNA     ***l*****q **t******* **e**g*qq* ***sv*sq** *****p**f*
Consensus  EGGVGFDYRL AMAIPDRWID YLKNKDDSEW SMSE-I--TL TNRRYTEKCI

           551                                                  600
RSBEI      ********** ********** ******t*** *********n **********
MSBEI      ********** ********** ******t*** ********** **********
D4cDNA     ********** *****m**** ******t*** ********** **********
PESBEII    s********* ********** **e***ss** c*tml***** ***s*h****
POSBE      ********** ********** ******s*** c*td***v** *****h****
D2cDNA     ****rqnh** **s**m**** **w*t*s*** a*d*d***** *a********
Consensus  AYAESHDQSI VGDKTIAFLL MDKEMY-GMS DLQPASPTID RGIALQKMIH
```

Figure 4 (cont..)

```
             601                                                    650
RSBEI        ********** ********** ********** ********** .*********
MSBEI        ********** ********** ********** ********** .*********
D4cDNA       ********** ********** ********** ********** .*****s*i*
PESBEII      ********** ********** ********** **g******* lt**n****n
POSBE        *f******** ********** ********** ********** .***n*a*s*
D2cDNA       *********s **k*****.. .......... .......... ..........
Consensus    FITMALGGDG YLNFMGNEFG HPEWIDFPRE GNNWSYDKCR -RQWSLVDTD

             651                                                    700
RSBEI        ********** *********e ********** ******k*** **********
MSBEI        ********** *********r ********** ********** **********
D4cDNA       ********** ********** ********** *******k** **********
PESBEII      ********** *r***l**** **i*a*t*** **st*n**** **********
POSBE        ********** *r***s**** ****a*g*** **s*d**n** **********
D2cDNA       .....***** v**vdtps** c******n*t a*h******g sa*tk*....
Consensus    HLRYKYMNAF DQAMNALD-K FSFLSSSKQI VSDMNEE-KV IVFERGDLVF

             701                                                    750
RSBEI        ******n*** k********* ********** **v******* **********
MSBEI        ******k*** ********** ********** **v******* **********
D4cDNA       ******s*** ********** ***k****** **m******* aqyn******
PESBEII      ******en** ********** ********** *te******* ***a*q****
POSBE        ******kn** ********** ********** *we******t **********
D2cDNA       .*thlrsgc* *p.....s** stssc**... .*gpsnqspf skpfig*pgc
Consensus    VFNFHP-KTY EGYKVGCDLP GKYRVALDSD AL-FGGHGRV GHDVDHFTSP

             751                                                    800
RSBEI        **m******* *****..... .......... .....***** **********
MSBEI        ********** *****..... .......... .....***** **********
D4cDNA       ********** *****..... .......... .....***** **********
PESBEII      ********** *****..... .......... .....***** ****h***v*
POSBE        ********** **g*qipskc cllrehvwli telmnacq*l kitrq*f*vs
D2cDNA       ifcc*lfkge *......... .......... .......... ..........
Consensus    EG-PGVPETN FNNRP----- ---------- -----NSFKV LSPPRTCVAY

             801                                                    850
RSBEI        *...****dr **l*rg**va s**i.vte** **e**s.... ..**ti**gw
MSBEI        *...****ag agr*lhak*e t***s**es* **k*s*.... ..a....ssk
D4cDNA       *...****ka *kpkde**** w**aa*g.** **e***vkda ad**at**sk
PESBEII      *...*****q **snnpnlg* *ee**a*adt **aripdvs* e*..ed*nld
POSBE        *yqqp*sr*v trnlkirylq *sv**tna*q klkf**qtf* v*yyqqpilr
D2cDNA       .......... .......... .......... .......... ..........
Consensus    Y---RVDER- EE-R--GAAS -GKT-PA-YI DV-ATR---- -SGE--SG--

             851                       876
RSBEI        kg***d*cg* **mk***r** *e*c*d
MSBEI        edk*atagg* **wk*arqp* *q*t**
D4cDNA       ka*tgg*ss* **in***g*p *k*n*.
PESBEII      r*e*ns**av dagi*kvere vvgdn*
POSBE        r*tr*lk*sl stnist*... ......
D2cDNA       .......... .......... ......
Consensus    --SEK-DD-K KG--FVF-SS D-D-K-
```

Figure 4 (cont..)

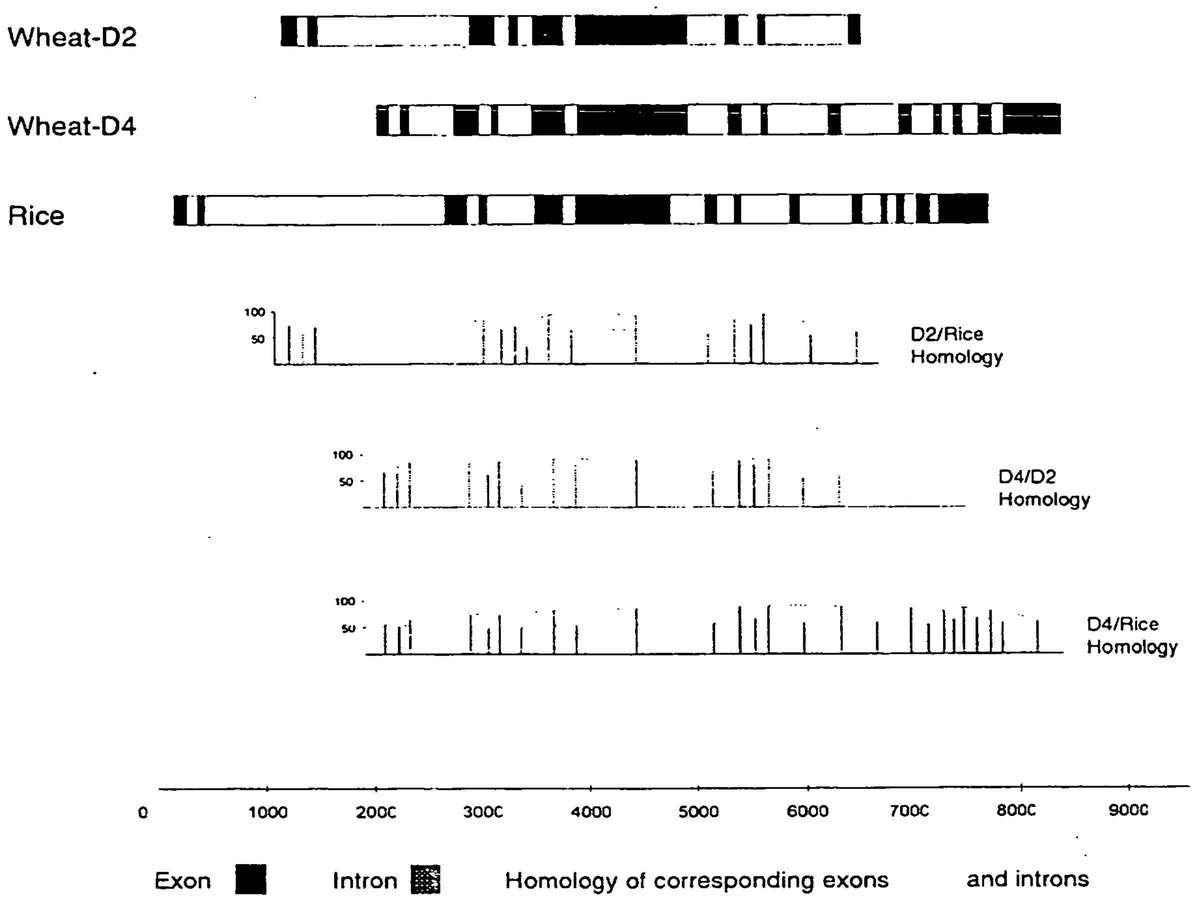

Wheat-D2
Wheat-D4
Rice
100
50
D2/Rice
Homology
100
50
D4/D2
Homology
100
50
D4/Rice
Homology
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Exon
Intron
Homology of corresponding exons and introns

FIGURE 5

DNA	5' TCCCGTGTCTGCGCCAAGAGACTACACCATGGCAACAGCTGAAGATGGTGTTGGCGACCT 3' 3' AGGGCACAGACGCGGTTCTCTGATGTGGTACCGTTGTCGACTTCTACCACAACCGCTGGA 5'
possible reading frames	S R V C A K R L H H G N S * R W C W R P P V S A P R D Y T M A T A E D G V G D L P C L R Q E T T P W Q Q L K M V L A T F
true N-terminal sequence for BE-1 (Morell et al, 1997)	V S A P R D Y T M A T A E D G V

Figure 6

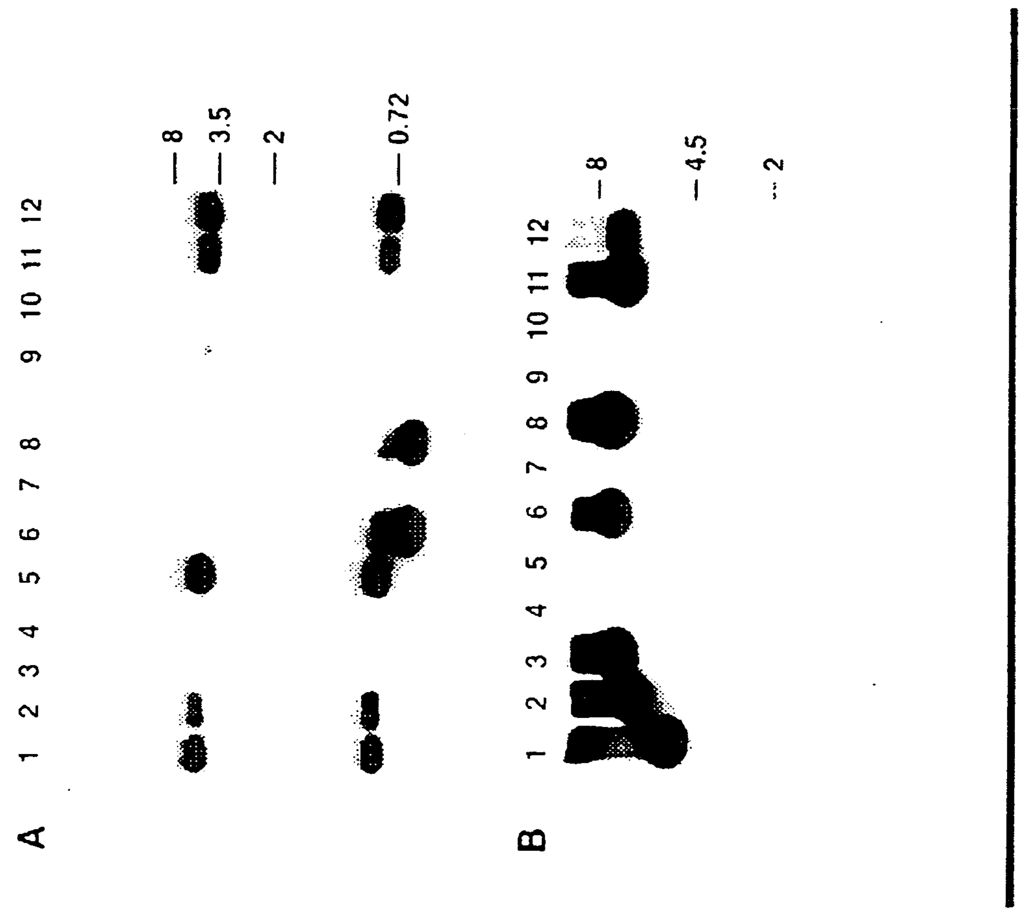


FIGURE 7

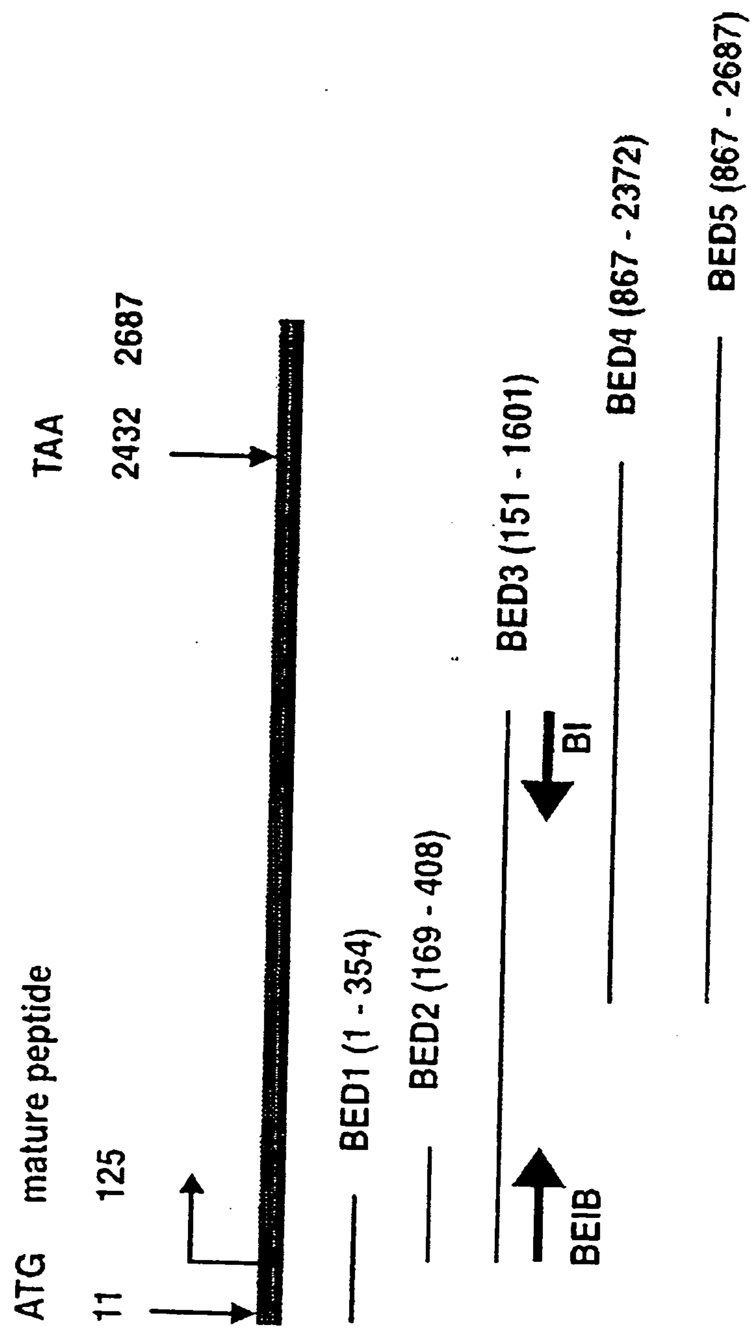

ATG
11
mature peptide
125
TAA
2432
2687
BED1 (1 - 354)
BED2 (169 - 408)
BED3 (151 - 1601)
BED4 (867 - 2372)
BED5 (867 - 2687)
BI
BEIB

FIGURE 8

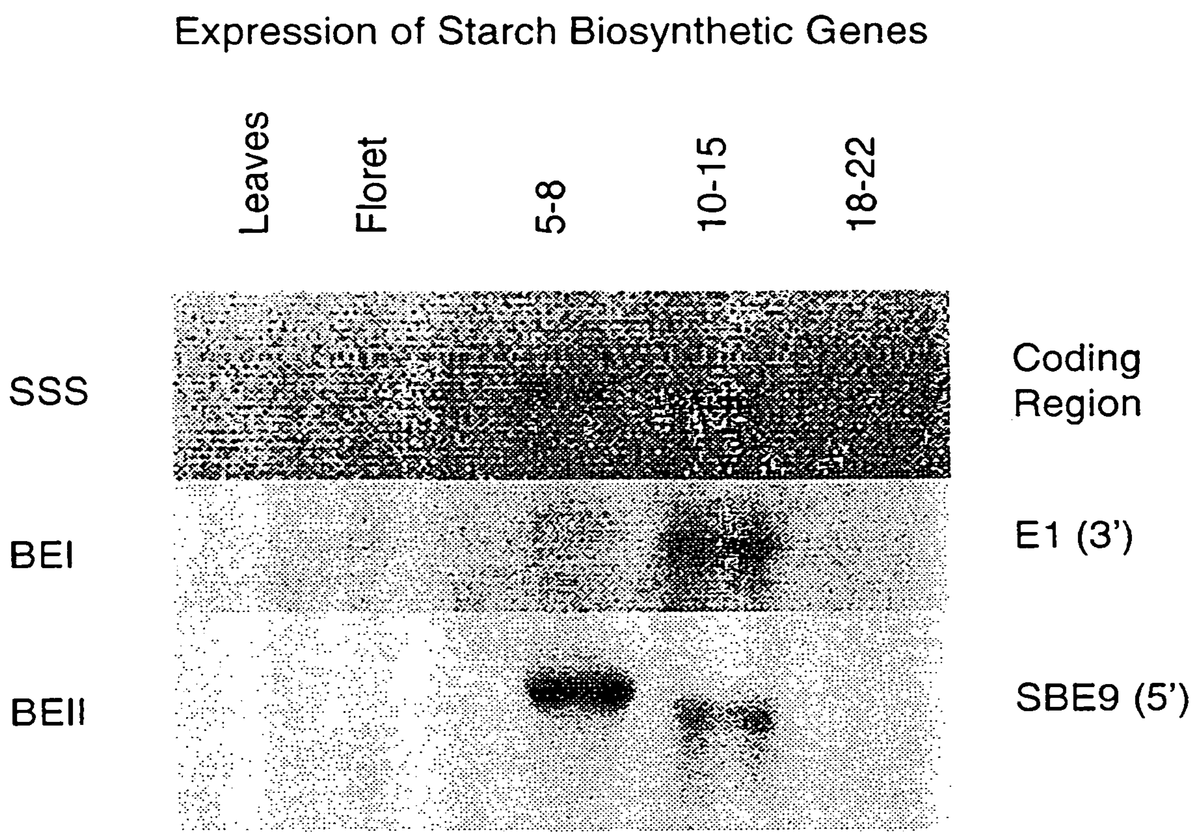

Expression of Starch Biosynthetic Genes
Leaves
Floret
5-8
10-15
18-22
SSS
BEI
BEII
Coding Region
E1 (3')
SBE9 (5')

FIGURE 9A

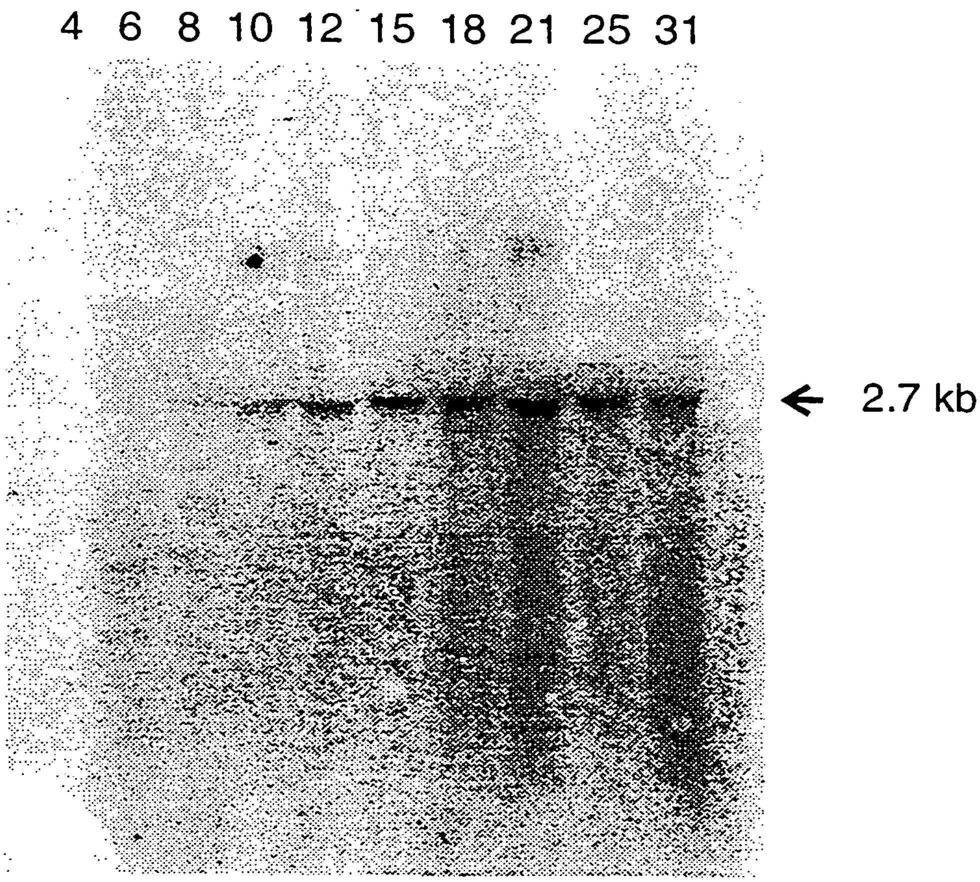


FIGURE 9B

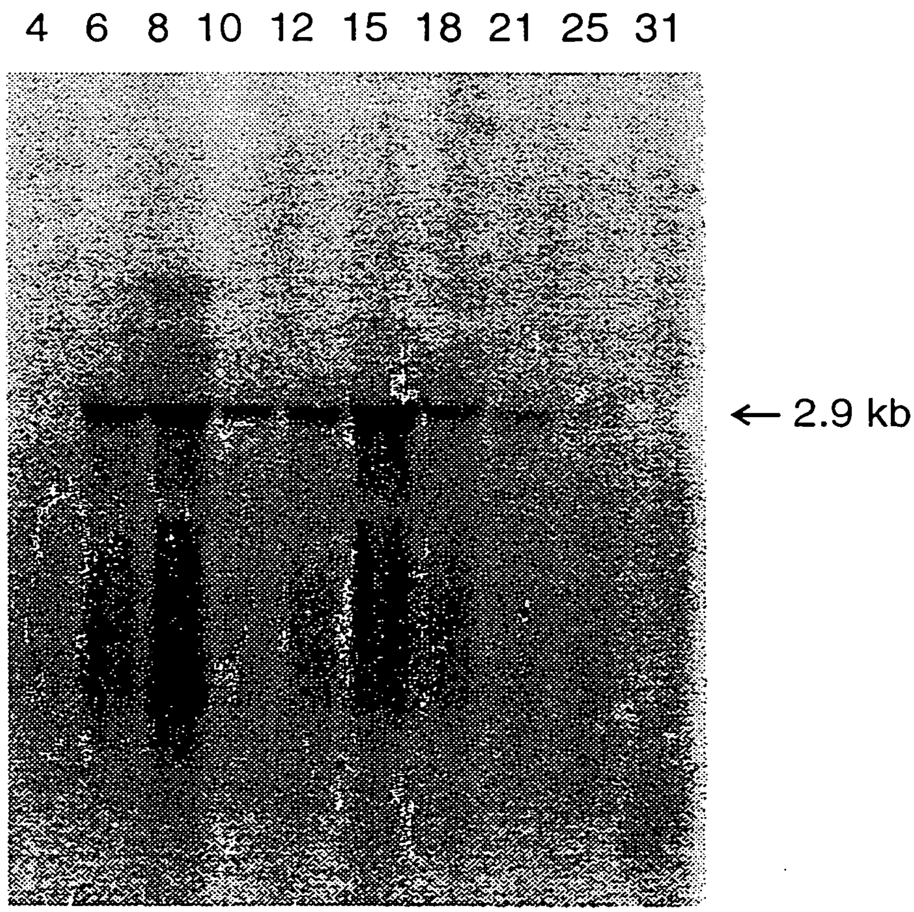


FIGURE 9C

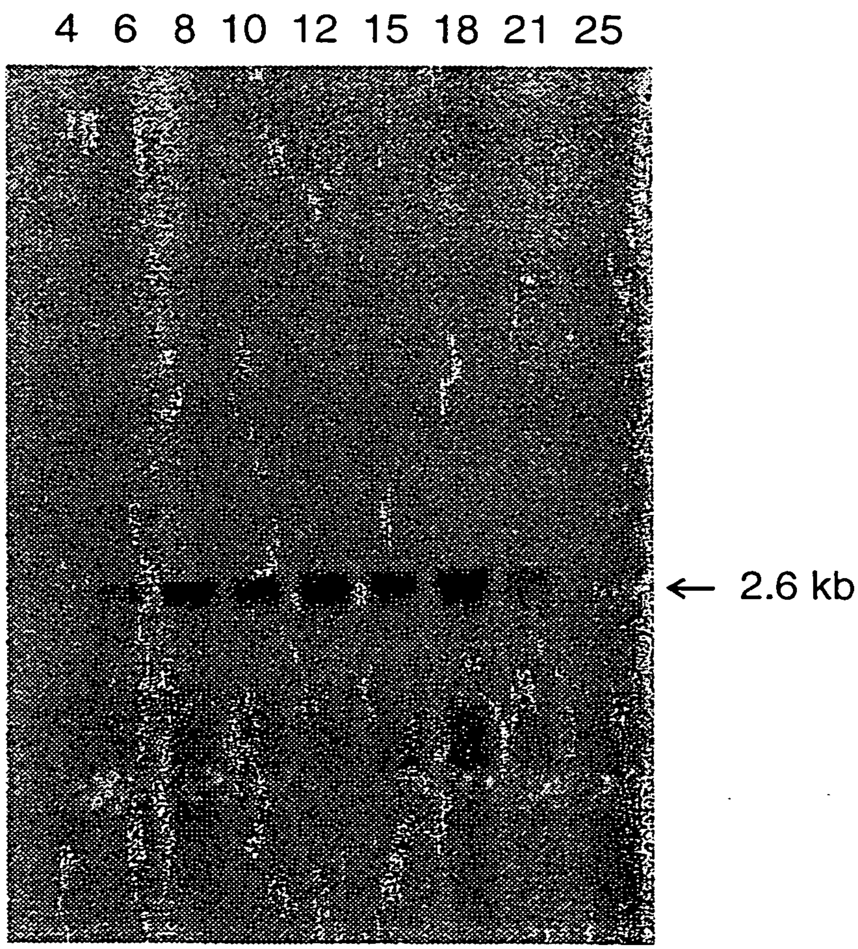


FIGURE 9D

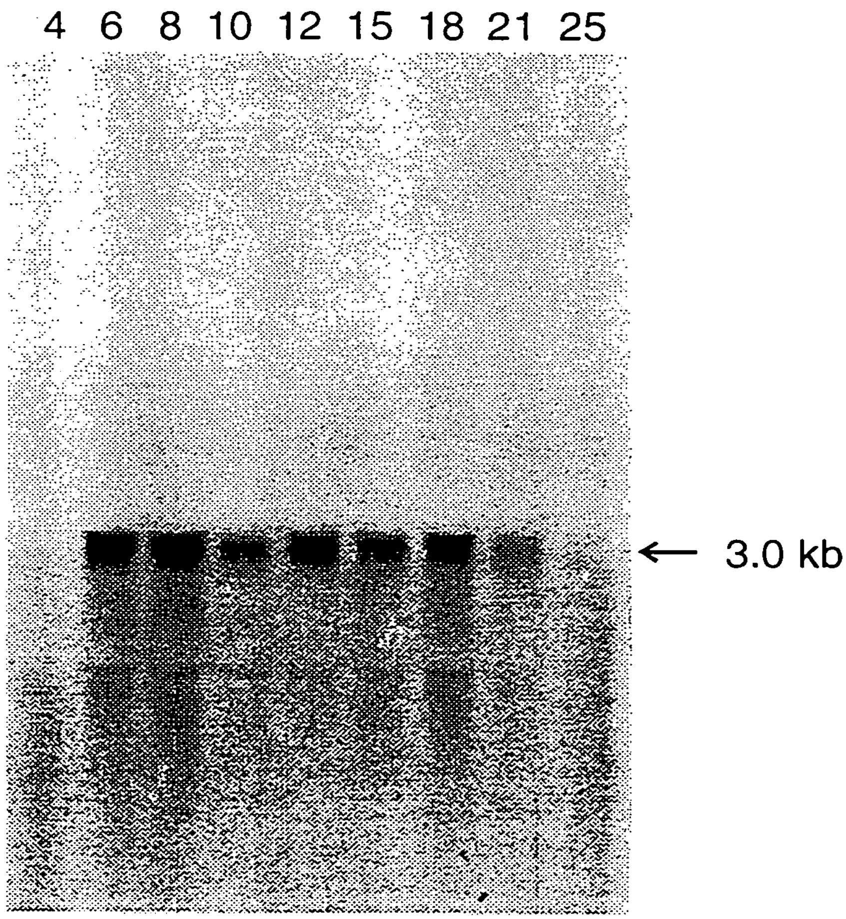


FIGURE 9E

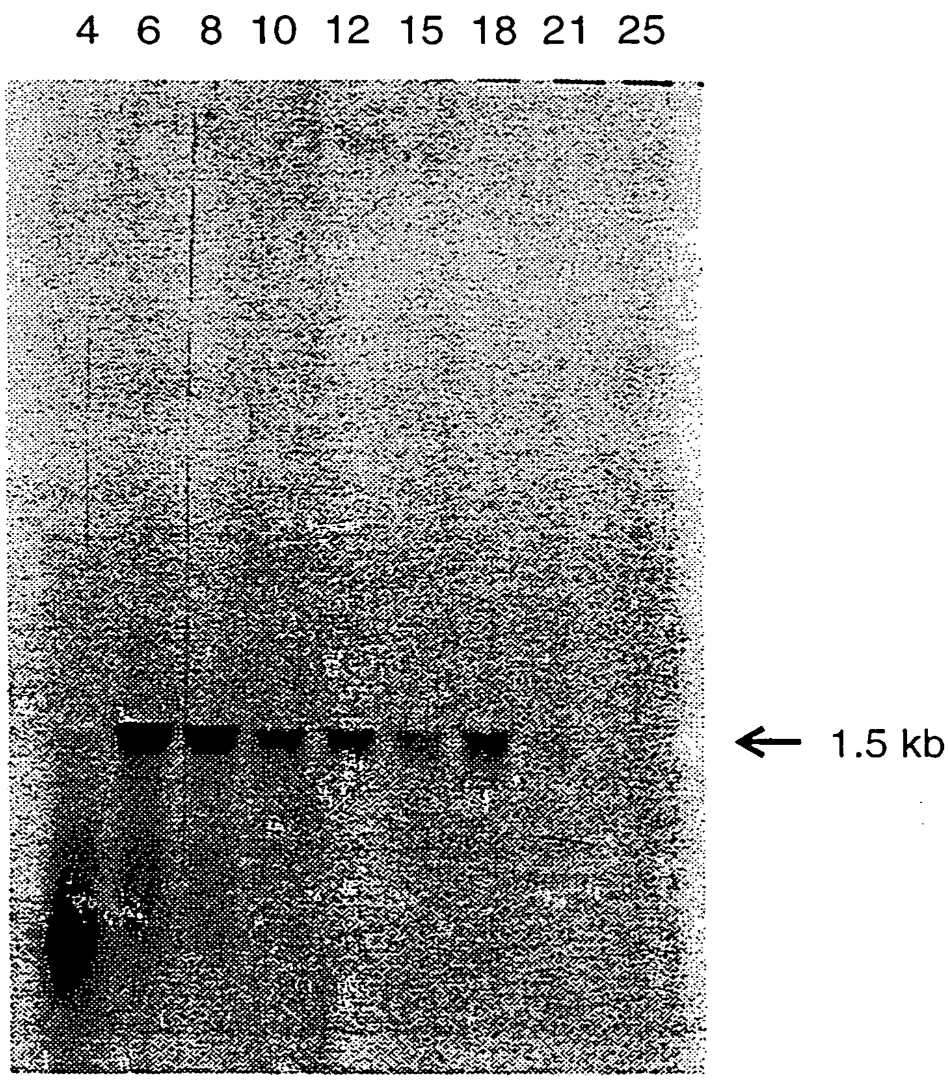


FIGURE 9F

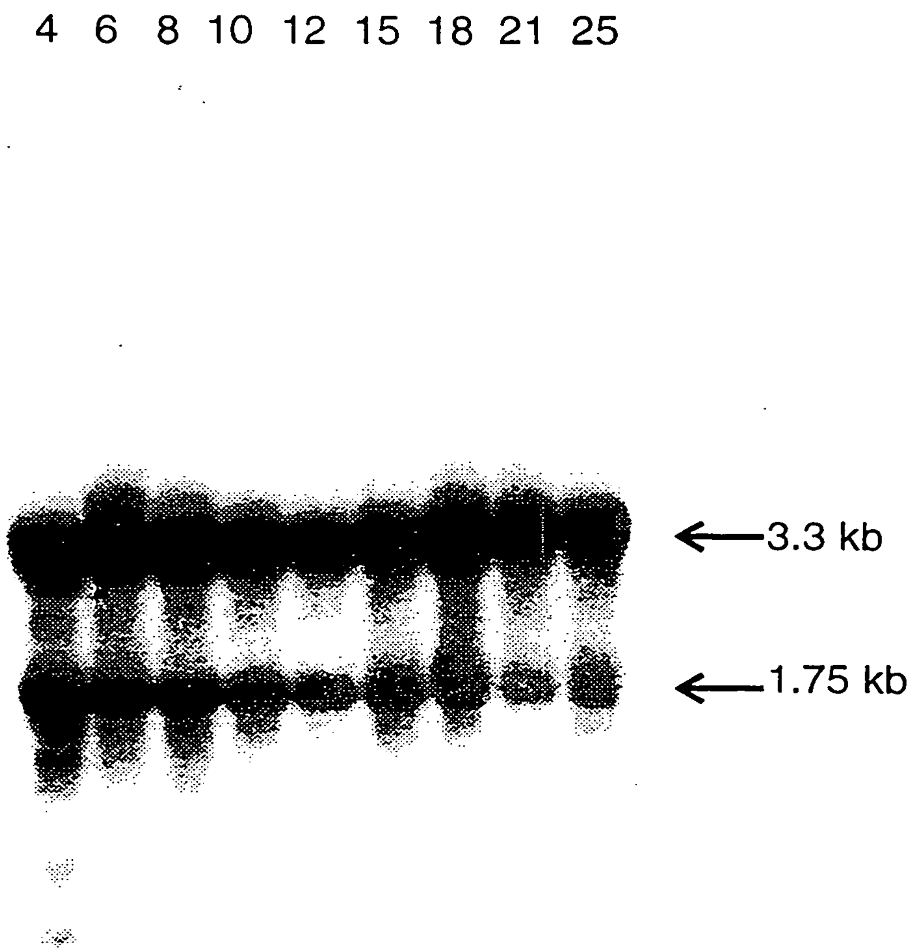


FIGURE 9G

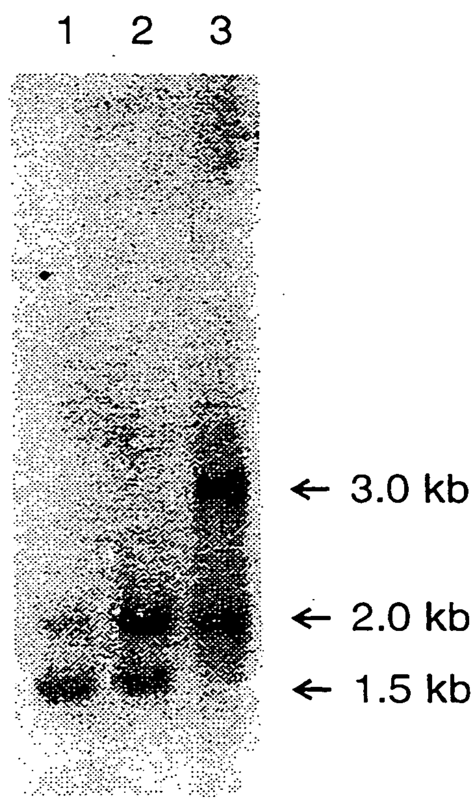


FIGURE 9H

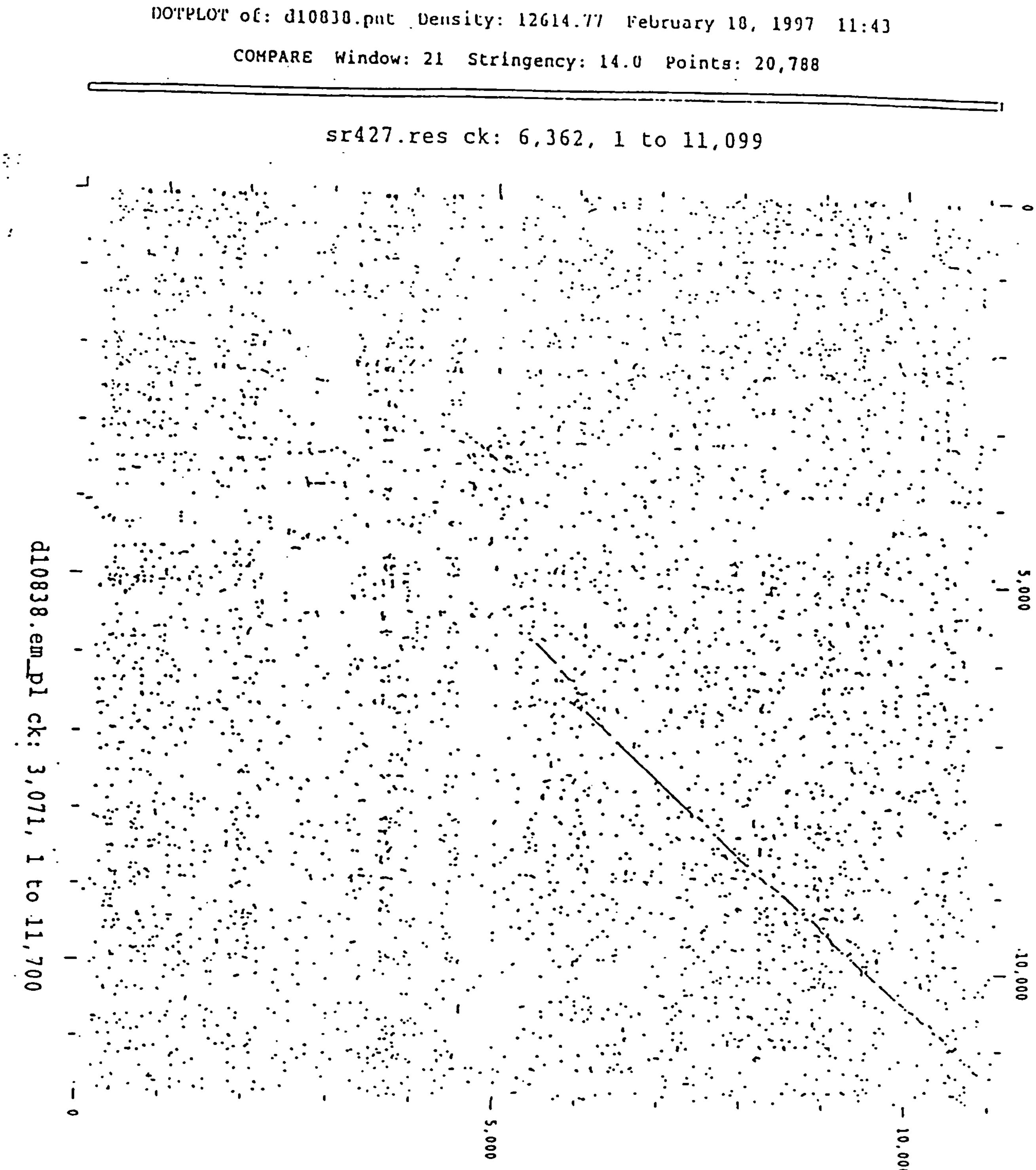

DOTPLOT of: d10838.pnt Density: 12614.77 February 18, 1997 11:43
COMPARE Window: 21 Stringency: 14.0 Points: 20,788
sr427.res ck: 6,362, 1 to 11,099
d10838.em_pl ck: 3,071, 1 to 11,700
0
5,000
10,000
0
5,000
10,000

Figure 10

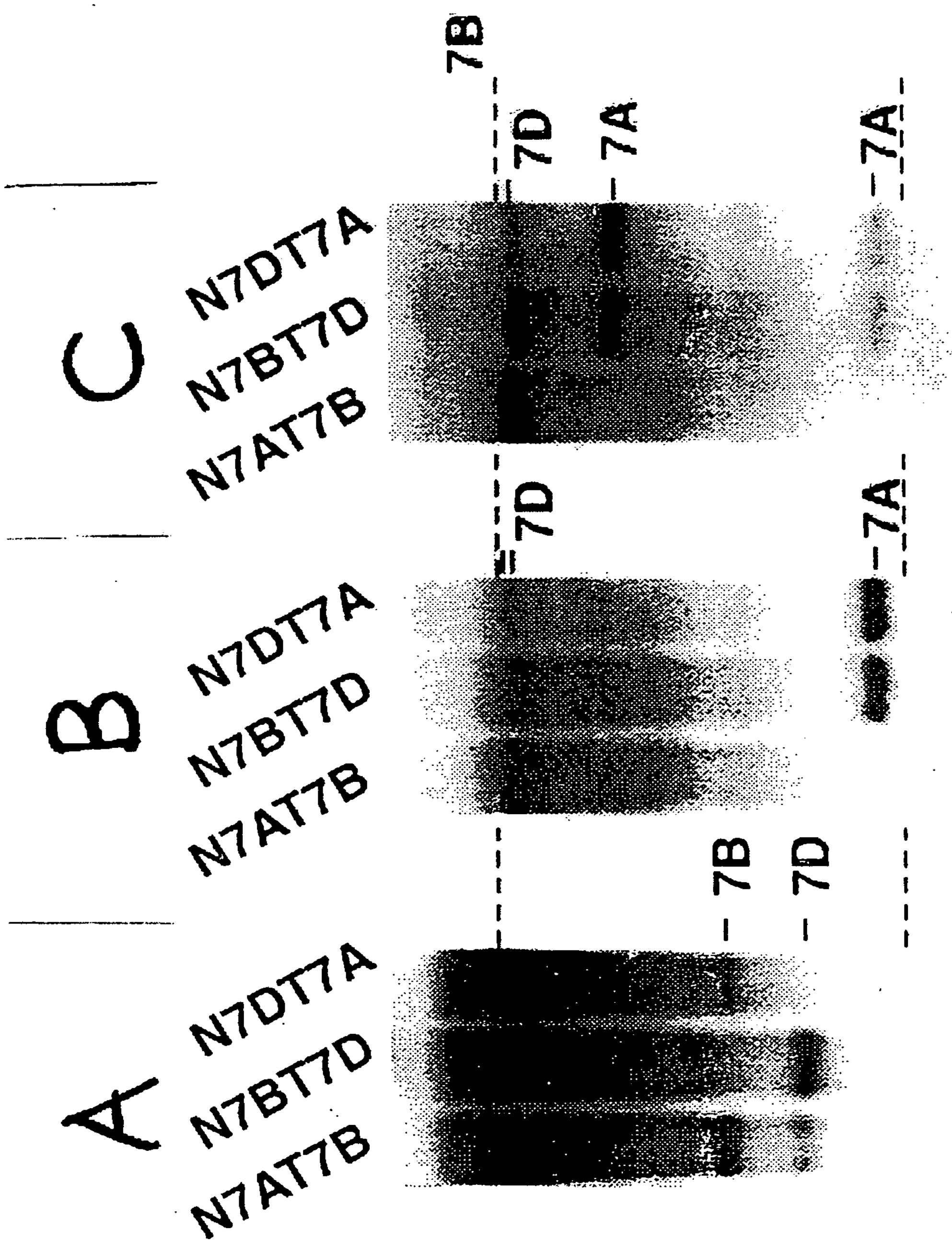

A
N7DT7A
N7BT7D
N7AT7B
7B
7D
B
N7DT7A
N7BT7D
N7AT7B
7D
7A
C
N7DT7A
N7BT7D
N7AT7B
7B
7D
7A
7A

FIGURE 11

Genomic Clones from *T.tauschii* for SBE II.

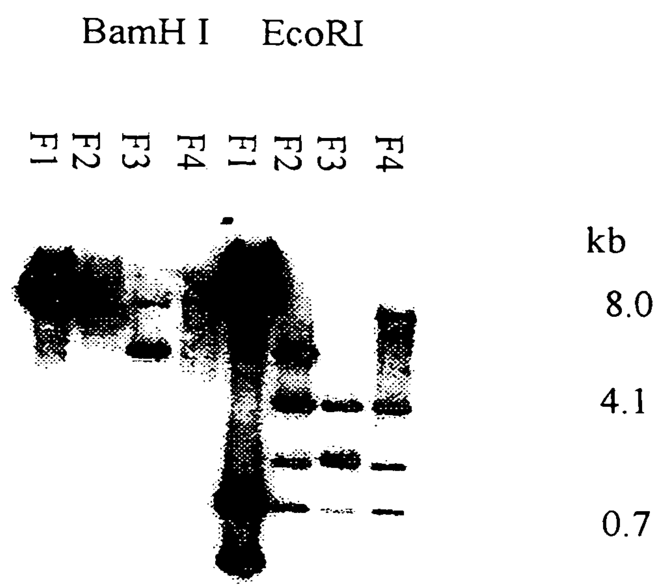


FIGURE 12

N-terminal sequences of cereal starch branching enzymes

Protein	1[A]	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
RICEBEI[B]	A	T	A	R	K	N	K	T	M	V	T	V	V	E	E	V						
WBE-I_{AD}	V	S	A	P	R	D	Y	T	M	A	T	A	E	D	G	V						
MAIZE BEI[C]	A	T	V	Q	E	D	K	T	M	A	T	A	K	G	D	V						
RICEBEII[D]	A	A	G	A	S	G	E	-	V	M	I	P	E	G	E	S	D	G	M	P	V	S
WBE-II		A	A	S	P	G	K	-	V	L	V	P	D	G	E	S	D	D	L	A	S	Y
MAIZE BEII[B]	A	A	A	A	A	R	K	A	V	M	V	P	E	G	E	N	D	G	L	A	S	

[A] N-terminal amino acid of the mature polypeptide. [B] Kawasaki *et al.*(1993), [C] Baba *et al.* (1991),

[D] Mizuno *et al.* (1993),[B] Fisher *et al.* (1993)

Residues in the wheat sequences showing identity with the respective maize or rice branching enzyme isoforms are highlighted in bold text.

Figure 13a

```
     TTCCCTTTTTTTTTCTTTGGGNGGGGGATGGCCTGTTGGATGNTGTTCCCCAATGAATTT
  1  ---------+---------+---------+---------+---------+---------+ 60
     AAGGGAAAAAAAAAGAAACCCNCCCCCTACCGGACAACCTACNACAAGGGGTTACTTAAA

a    F  P  F  F  F  F  G  ?  G  M  A  C  W  M  ?  F  P  N  E  F  -
b     S  L  F  F  S  L  G  G  G  W  P  V  G  ?  C  S  P  M  N  F -
c      P  F  F  F  L  W  ?  G  D  G  L  L  D  ?  V  P  Q  *  I  S-

     CCATGGAGTGAGAGAGATAGTTGGATNAGGGATCGCGNTTCCNGGAACTGTATTTTTTTC
 61  ---------+---------+---------+---------+---------+---------+ 120
     GGTACCTCACTCTCTCTATCAACCTANTCCCTAGCGCNAAGGNCCTTGACATAAAAAAAG

a    P  W  S  E  R  D  S  W  ?  R  D  R  ?  S  ?  N  C  I  F  F  -
b     H  G  V  R  E  I  V  G  ?  G  I  A  ?  P  G  T  V  F  F  S -
c      M  E  *  E  R  *  L  D  ?  G  S  R  F  ?  E  L  Y  F  F  P-

     CCCNGCGGGGGAAATGGCGTTAGTGTCNACCCAGGCCCTGGTGTTACCACGGCTTTGATC
121  ---------+---------+---------+---------+---------+---------+ 180
     GGGNCGCCCCCTTTACCGCAATCACAGNTGGGTCCGGGACCACAATGGTGCCGAAACTAG

a    P  ?  G  G  N  G  V  S  V  ?  P  G  P  G  V  T  T  A  L  I  -
b     P  A  G  E  M  A  L  V  S  T  Q  A  L  V  L  P  R  L  *  S -
c      ?  R  G  K  W  R  *  C  ?  P  R  P  W  C  Y  H  G  F  D  H-

     ATTCTTCGTTTCATTCTGATATATATTTTCTCATTCTTTTTCTTCCTGTTCTTGCTGTAA
181  ---------+---------+---------+---------+---------+---------+ 240
     TAAGAAGCAAAGTAAGACTATATATAAAAGAGTAAGAAAAAGAAGGACAAGAACGACATT

a    I  L  R  F  I  L  I  Y  I  F  S  F  F  F  F  L  F  L  L  *  -
b     F  F  V  S  F  *  Y  I  F  S  H  S  F  S  S  C  S  C  C  N -
c      S  S  F  H  S  D  I  Y  F  L  I  L  F  L  P  V  L  A  V  T-

     CTGCAAGTTGTGGCGTTTTTTCACTATTGTAGTCATCCTTGCATTTTGCAGGCGCCGTCC
241  ---------+---------+---------+---------+---------+---------+ 300
     GACGTTCAACACCGCAAAAAAGTGATAACATCAGTAGGAACGTAAAACGTCCGCGGCAGG

a    L  Q  V  V  A  F  F  H  Y  C  S  H  P  C  I  L  Q  A  P  S  -
b     C  K  L  W  R  F  F  T  I  V  V  I  L  A  F  C  R  R  R  P -
c      A  S  C  G  V  F  S  L  L  *  S  S  L  H  F  A  G  A  V  L-

     TGAGCCGGCGGCGCCTCTCCAGGGAAGGTCCTGGTGCCTGACGGCGAGAGNGACGACTTGG
301  ---------+---------+---------+---------+---------+---------+ 360
     ACTCGGCGCGCCGGAGAGGTCCCTTCCAGGACCACGGACTGCCGCTCTCNCTGCTGAACC

a    *  A  A  R  P  L  Q  G  R  S  W  C  L  T  A  R  ?  T  T  W  -
b     E  P  R  G  L  S  R  E  G  P  G  A  *  R  R  E  ?  R  L  G -
c      S  R  A  A  S  P  G  K  V  L  V  P  D  G  E  ?  D  D  L  A-

     CAAGTCCGGCGCAACCTGAAGAATTACAGGTACACACACTCGTGCCGGTAAATCTTCATA
361  ---------+---------+---------+---------+---------+---------+ 420
     GTTCAGGCCGCGTTGGACTTCTTAATGTCCATGTGTGTGAGCACGGCCATTTAGAAGTAT

a    Q  V  R  R  N  L  K  N  Y  R  Y  T  H  S  C  R  *  I  F  I  -
b     K  S  G  A  T  *  R  I  T  G  T  H  T  R  A  G  K  S  S  Y -
c      S  P  A  Q  P  E  E  L  Q  V  H  T  L  V  P  V  N  L  H  T-

     CAATCGTTATTCACTTACCAAATGCCGGATGAAACCAACCACGGATGCGTCAGGTTTCGA
421  ---------+---------+---------+---------+---------+---------+ 480
     GTTAGCAATAAGTGAATGGTTTACGGCCTACTTTGGTTGGTGCCTACGCAGTCCAAAGCT

a    Q  S  L  F  T  Y  Q  M  P  D  E  T  N  H  G  C  V  R  F  R  -
b     N  R  Y  S  L  T  K  C  R  M  K  P  T  T  D  A  S  G  F  E -
```

Figure 13b

Branching Enzyme-II Genes

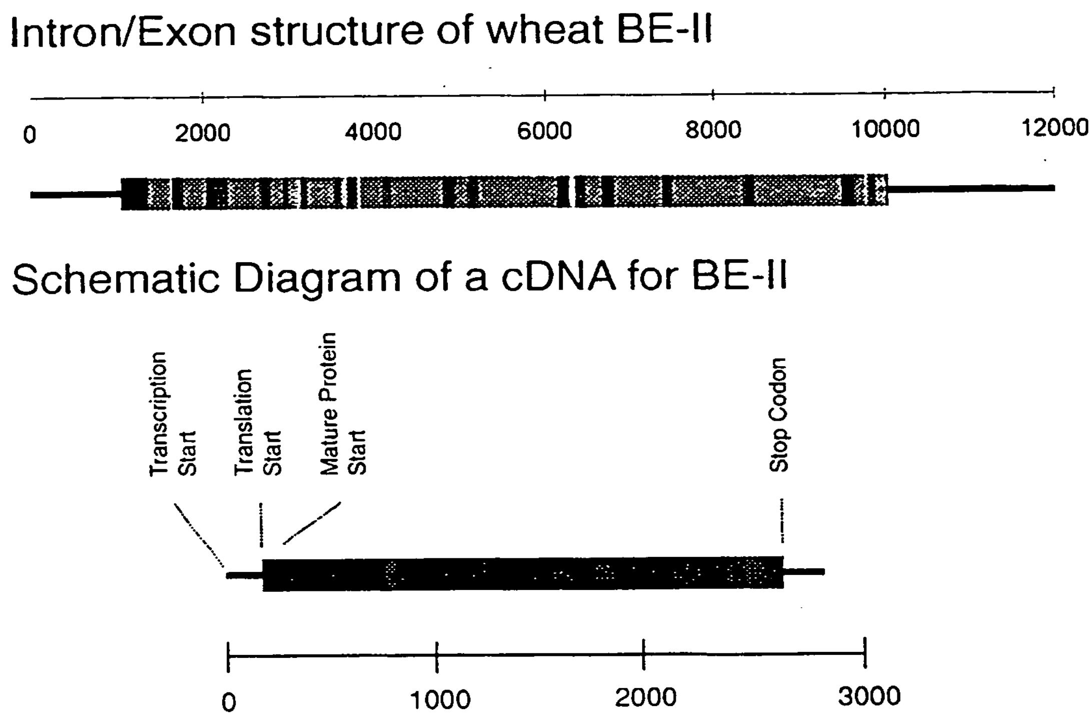


FIGURE 14

Wheat DNA probed with the
5' conserved sequence of SBE II.

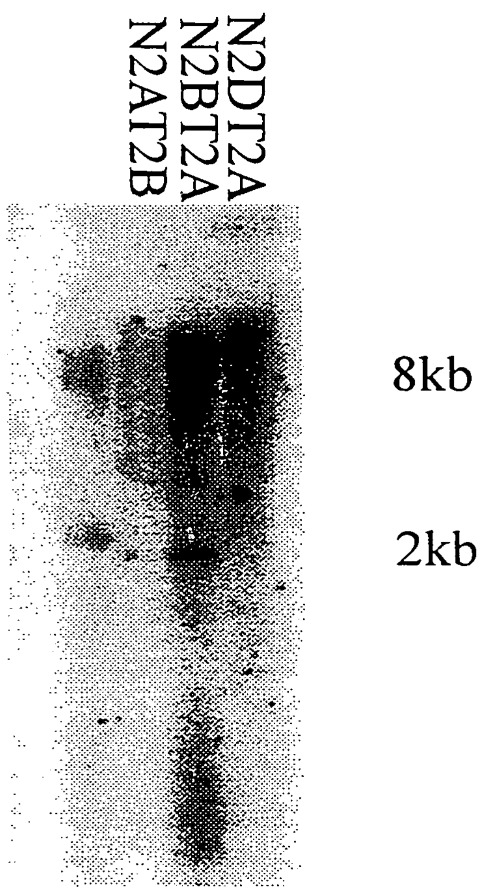


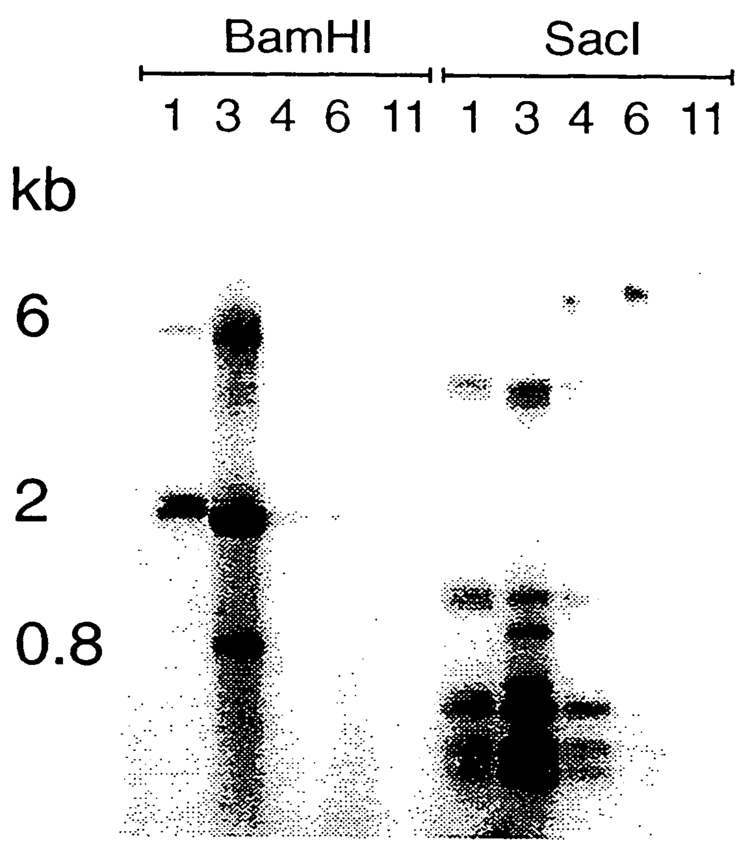
FIGURE 15

COMPARISON OF N-TERMINAL SEQUENCES
OF SOLUBLE STARCH SYNTHASE

GRYVAELSREGPAAARP	Deduced from wheat cDNA
GPYVAELSPEGPAAAPP	Wheat N-terminal

Figure 16

Soluble Starch Synthase Genomic Clones



Probed with SM-2 full length cDNA

FIGURE 17

INTRON EXON STRUCTURE - Wheat SSI

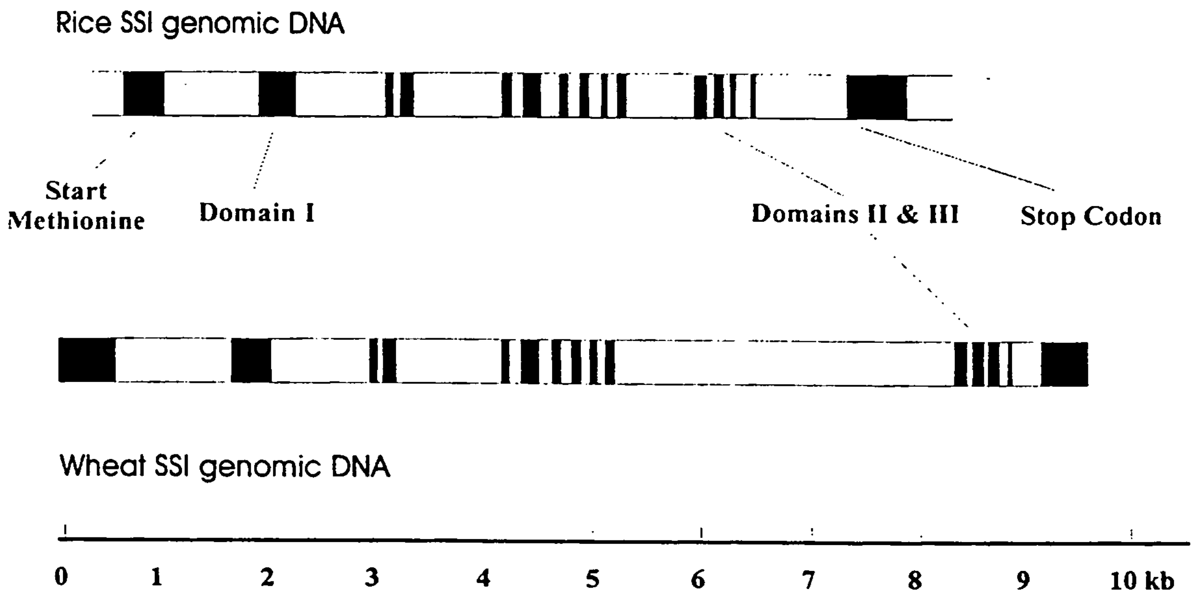


FIGURE 18

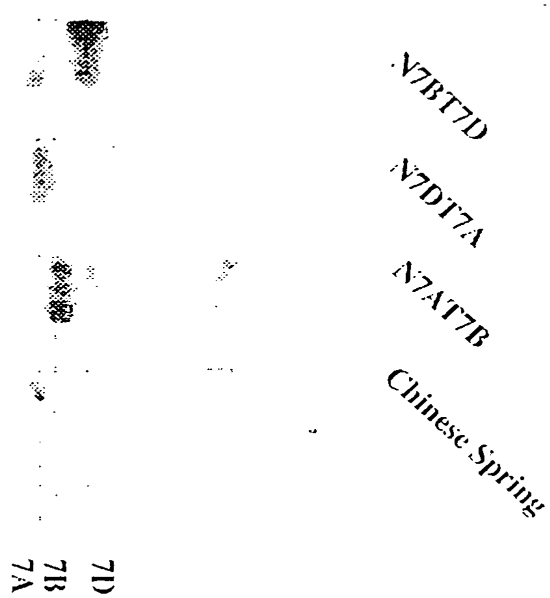


FIGURE 19

```
       ATACTACATACTATATGCTTGCACCCAAGGGACACTTTTATAACTATTCTGGCTGTGGGA
    80 +---------+---------+---------+---------+---------+--------- 139
       TATGATGTATGATATACGAACGTGGGTTCCCTGTGAAAATATTGATAAGACCGACACCCT

a        T  T  Y  Y  M  L  A  P  K  G  H  F  Y  N  Y  S  G  C  G  N -
b       I  L  H  T  I  C  L  H  P  R  D  T  F  I  T  I  L  A  V  G  -
c        Y  Y  I  L  Y  A  C  T  Q  G  T  L  L  *  L  F  W  L  W  E -

       ATACCTTCAACTGTAATCATCCTGTGGTTCGTCAATTCATTGTAGATTGTTTAAGATACT
   140 +---------+---------+---------+---------+---------+--------- 199
       TATGGAAGTTGACATTAGTAGGACACCAAGCAGTTAAGTAACATCTAACAAATTCTATGA

a        T  F  N  C  N  H  P  V  V  R  Q  F  I  V  D  C  L  R  Y  W -
b       I  P  S  T  V  I  I  L  W  F  V  N  S  L  *  I  V  *  D  T  -
c        Y  L  Q  L  *  S  S  C  G  S  S  I  H  C  R  L  F  K  I  L -

       GGGTGACGGAAATGCATGTTGATGGTTTTCGTTTTGACCTT
   200 +---------+---------+---------+---------+ 240
       CCCACTGCCTTTACGTACAACTACCAAAAGCAAAACTGGAA

a        V  T  E  M  H  V  D  G  F  R  F  D  L   -
b       G  *  R  K  C  M  L  M  V  F  V  L  T    -
c        G  D  G  N  A  C  *  W  F  S  F  *  P   -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

EcoRI

Figure 20a

Comparison of Wheat Debranching Enzyme-I (WDBE-I) PCR fragment with maize *Sugary*-1 DNA sequence

```
                  1098     1107      1117      1127      1137      1147      1157
SUGARY.DNA         TGAGGTGATCATGGATGTTGTCTTCAATCATACAGCTGAAGGTAATGAGAAAGGCCCAAT
                       ||||||||||||||||||||||| ||||||||||| ||||||||||| || |||||
WHEAT1.DNA         ....GTGATCATGGATGTTGTCTTCAACCATACAGCTGAGGGTAATGAGAATGGTCCAAT
                   -3       6         16        26        36        46        56

FILE NAME         1158     1167      1177      1187      1197      1207      1217
SUGARY.DNA         ATTATCCTTTAGGGGGATAGATAATAGTACATACTACATGCTTGCACCTAAGGGAGAGTT
                   |||||| ||||||||| | ||||||| ||||||||| ||||||||||| |||||| | ||
WHEAT1.DNA         ATTATCATTTAGGGGGGTCGATAATACTACATACTATATGCTTGCACCCAAGGGACACTT
                   57       66        76        86        96        106       116

FILE NAME         1218     1227      1237      1247      1257      1267      1277
SUGARY.DNA         TTATAATTATTCTGGTTGTGGAAATACCTTCAATTGTAATCATCCTGTAGTCCGTGAATT
                   |||||| |||||||| |||||  |||||||||| |||||||||||||| || ||| ||||
WHEAT1.DNA         TTATAACTATTCTGGCTGTGGGNATACCTTCAACTGTAATCATCCTGTGGTTCGTCAATT
                   117      126       136       146       156       166       176

FILE NAME         1278     1287      1297      1307      1317      1327      1337
SUGARY.DNA         TATAGTGGATTGCTTGAGATACTGGGTAACAGAAATGCATGTTGATGGTTTTCGTTTTGA
                    || || ||||| || || |||||||| || ||||||||||||| |||||||||||||||
WHEAT1.DNA         CATTGTAGATTGTTTAAGNTACTGGGTGACGGAAATGCATGTTGNTGGTTTTCGTTTTGA
                   177      186       196       206       216       226       236

FILE NAME         1338     1347      1357
SUGARY.DNA         CCTTGCATCTATACT-G...
                   ||||||||||   ||
WHEAT1.DNA         CCTTGCATCTN--CTTNAAA
                   237      246       256

MATCHING PERCENTAGE
      TOTAL WINDOW        84%     (  219/  260)
      ALIGNMENT WINDOW    86%     (  219/  253)
```

Figure 20b

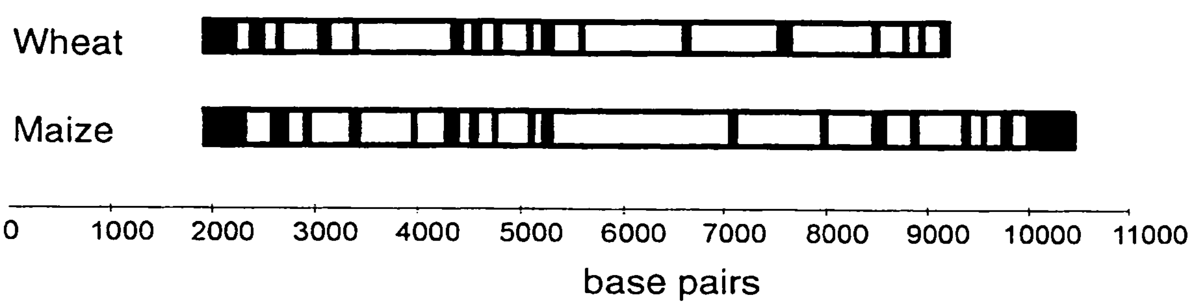

Wheat
Maize
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
base pairs

FIGURE 20C

Southern blot of T. *tauschii*
Genomic DNA

1X 3X

BamHI Digest

T. tauschii Genomic DNA Probed
With The Wheat Debranching Enzyme
PCR Product

FIGURE 21A

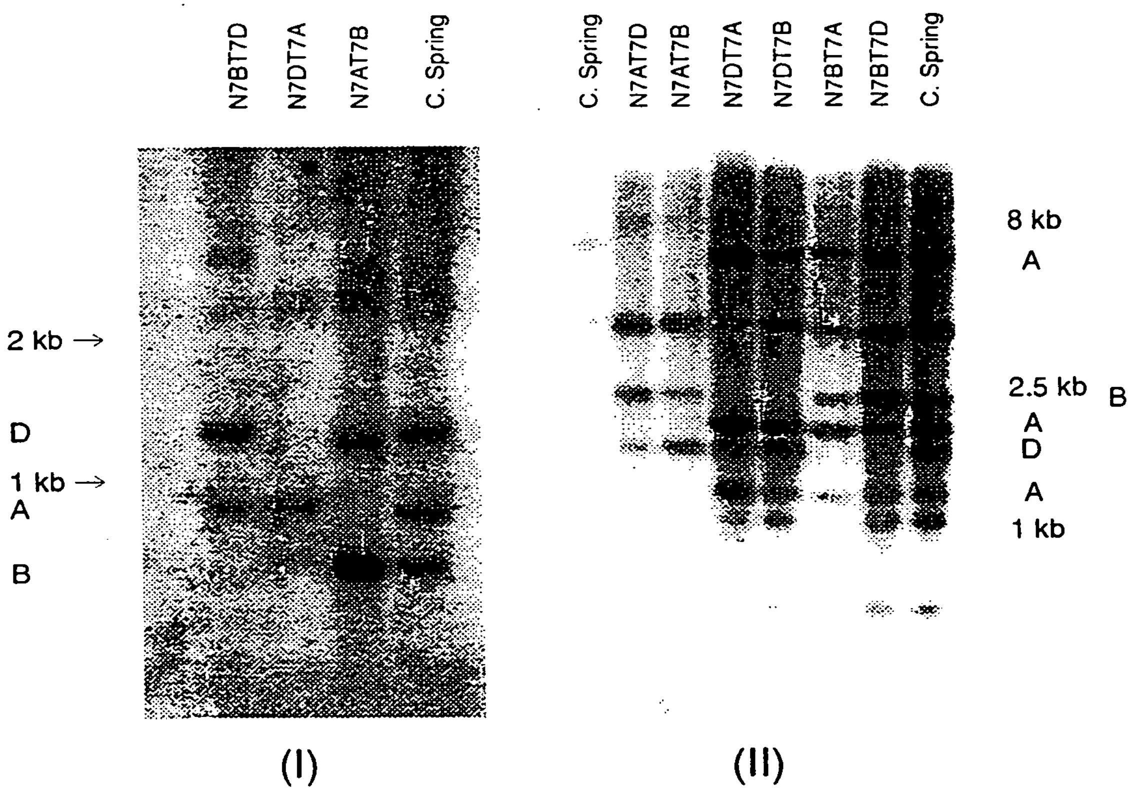

N7BT7D
N7DT7A
N7AT7B
C. Spring
2 kb →
D
1 kb →
A
B
(I)
C. Spring
N7AT7D
N7AT7B
N7DT7A
N7DT7B
N7BT7A
N7BT7D
C. Spring
8 kb
A
2.5 kb
B
A
D
A
1 kb
(II)

FIGURE 21B

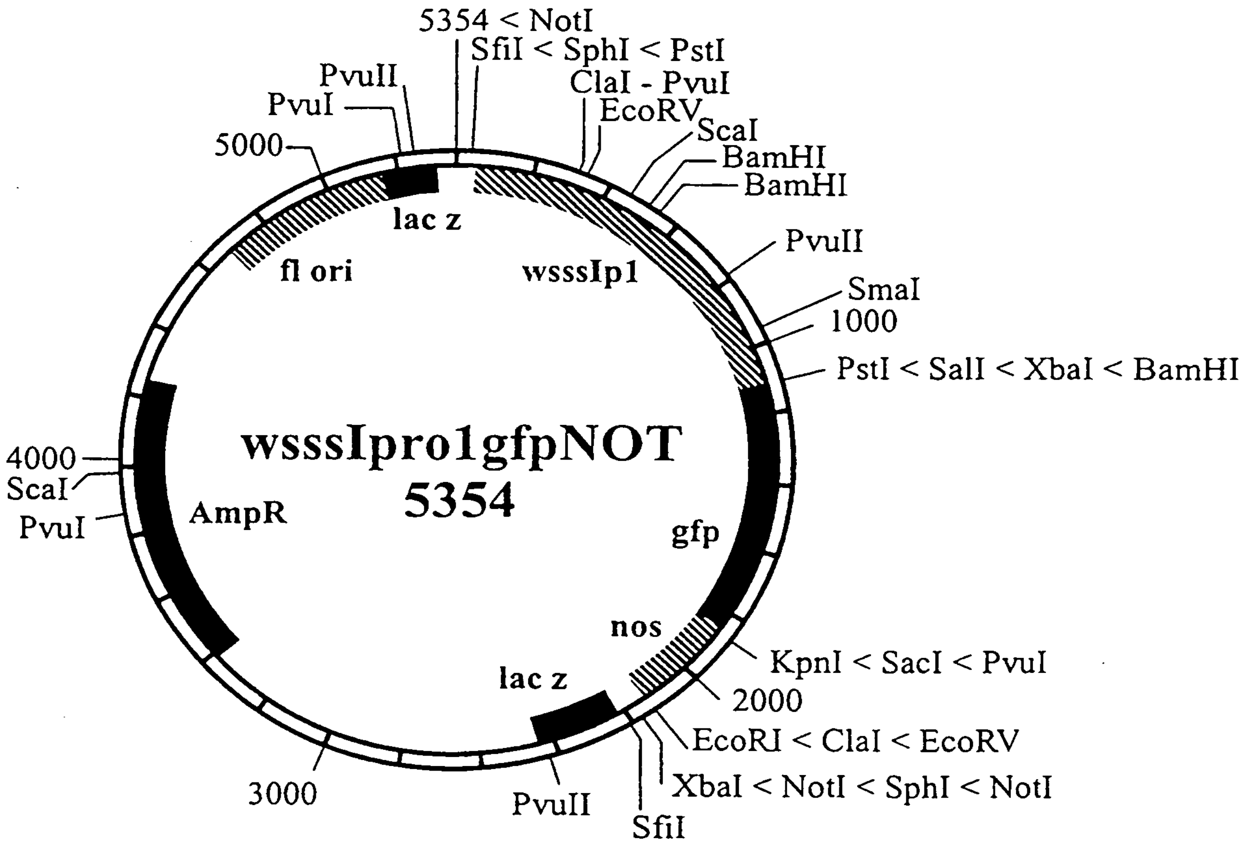

5354 < NotI
SfiI < SphI < PstI
ClaI - PvuI
EcoRV
ScaI
BamHI
BamHI
PvuII
SmaI
1000
PstI < SalI < XbaI < BamHI
KpnI < SacI < PvuI
2000
EcoRI < ClaI < EcoRV
XbaI < NotI < SphI < NotI
SfiI
PvuII
3000
PvuI
ScaI
4000
5000
PvuI
PvuII
lac z
fl ori
wsssIp1
wsssIpro1gfpNOT
5354
AmpR
gfp
nos
lac z

FIGURE 22A

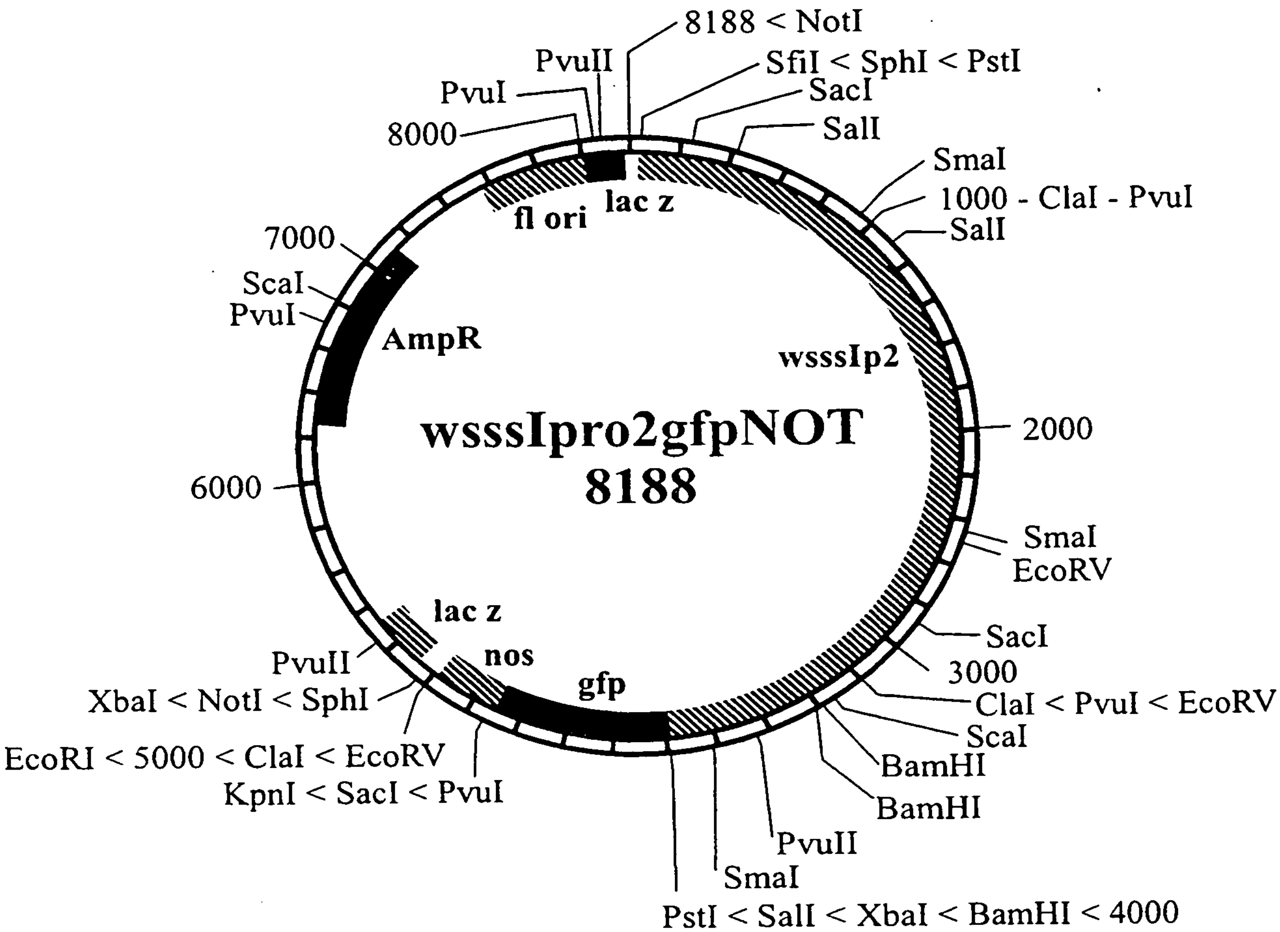

8188 < NotI
SfiI < SphI < PstI
SacI
SalI
SmaI
1000 - ClaI - PvuI
SalI
wsssIp2
2000
SmaI
EcoRV
SacI
3000
ClaI < PvuI < EcoRV
ScaI
BamHI
BamHI
PvuII
SmaI
PstI < SalI < XbaI < BamHI < 4000
gfp
nos
lac z
KpnI < SacI < PvuI
EcoRI < 5000 < ClaI < EcoRV
XbaI < NotI < SphI
PvuII
6000
AmpR
PvuI
ScaI
7000
fl ori
lac z
8000
PvuI
PvuII
wsssIpro2gfpNOT
8188

FIGURE 22B

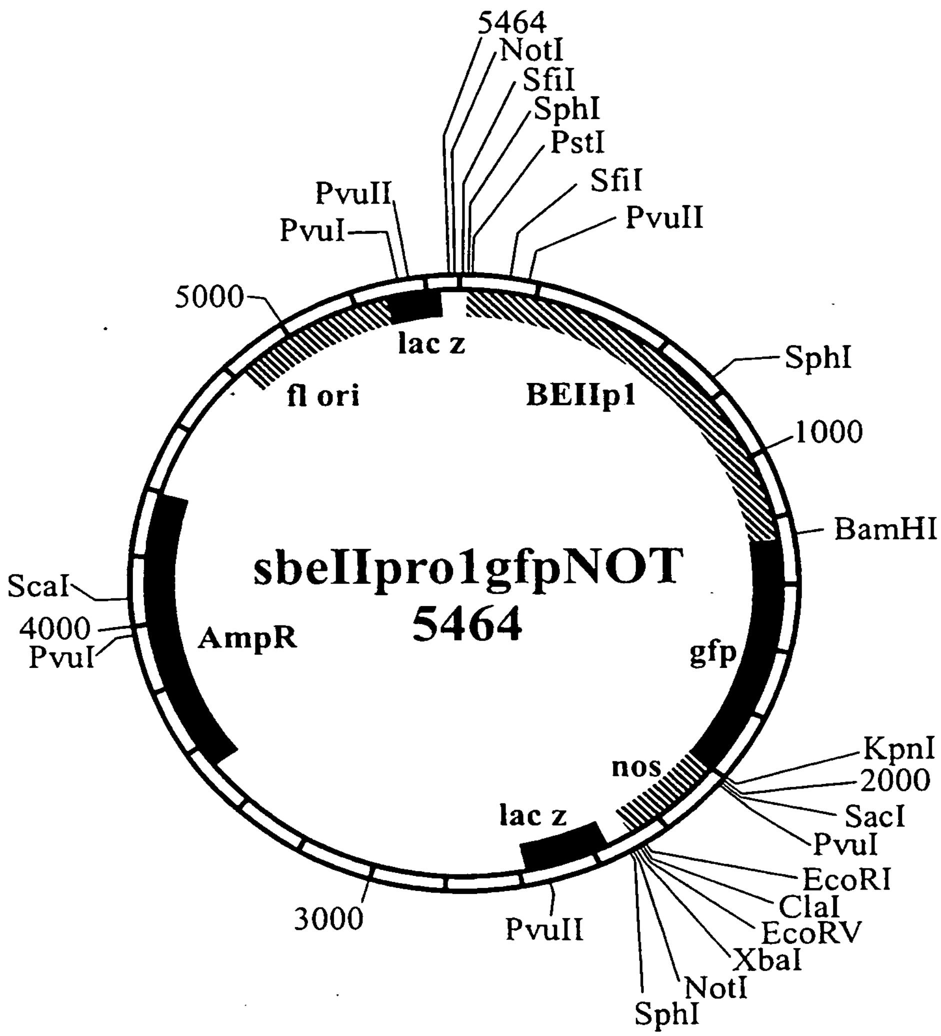

5464
NotI
SfiI
SphI
PstI
SfiI
PvuII
PvuII
PvuI
5000
lac z
fl ori
BEIIp1
SphI
1000
BamHI
sbeIIpro1gfpNOT
5464
ScaI
4000
PvuI
AmpR
gfp
KpnI
2000
SacI
PvuI
nos
lac z
EcoRI
ClaI
EcoRV
XbaI
NotI
SphI
3000
PvuII

FIGURE 22C

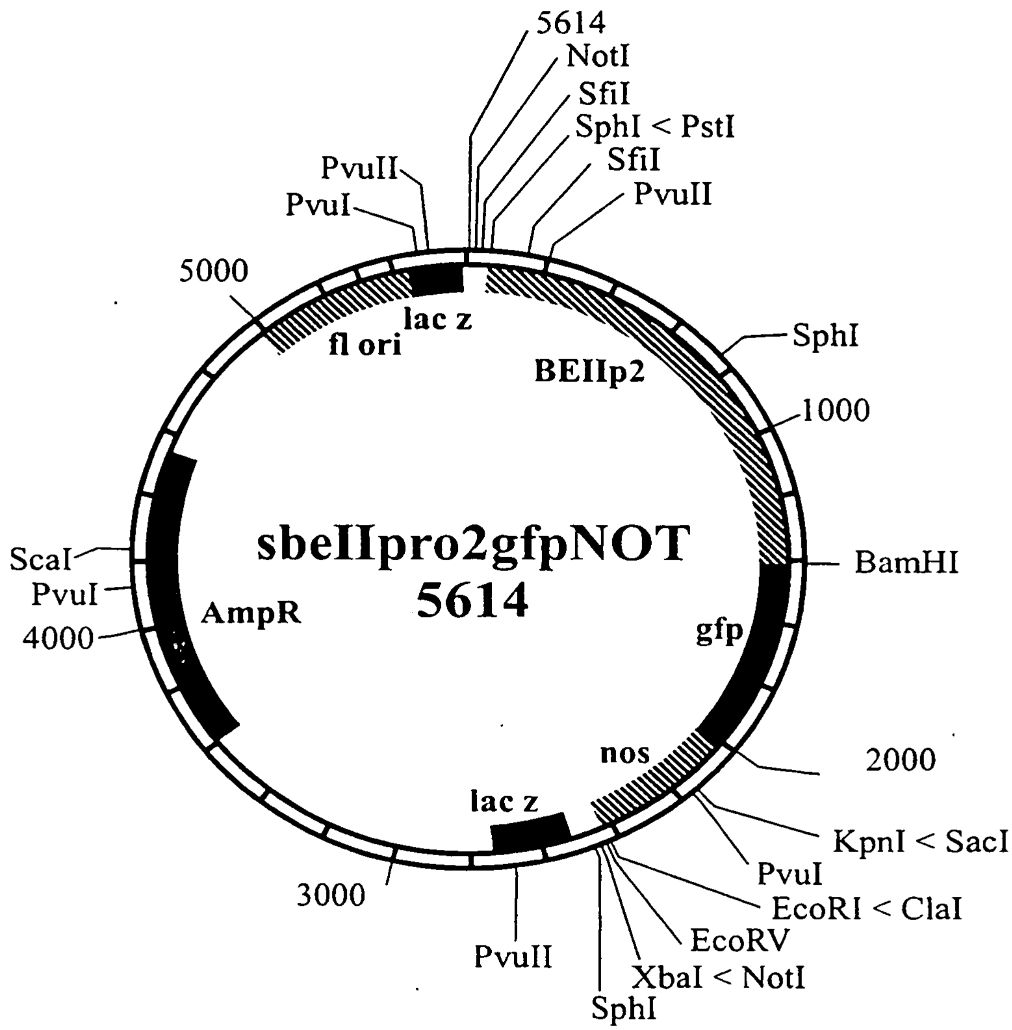

5614
NotI
SfiI
SphI < PstI
SfiI
PvuII
PvuII
PvuI
5000
lac z
fl ori
BEIIp2
SphI
1000
sbeIIpro2gfpNOT
5614
ScaI
PvuI
4000
AmpR
BamHI
gfp
nos
2000
lac z
KpnI < SacI
PvuI
EcoRI < ClaI
EcoRV
XbaI < NotI
SphI
3000
PvuII

FIGURE 22D

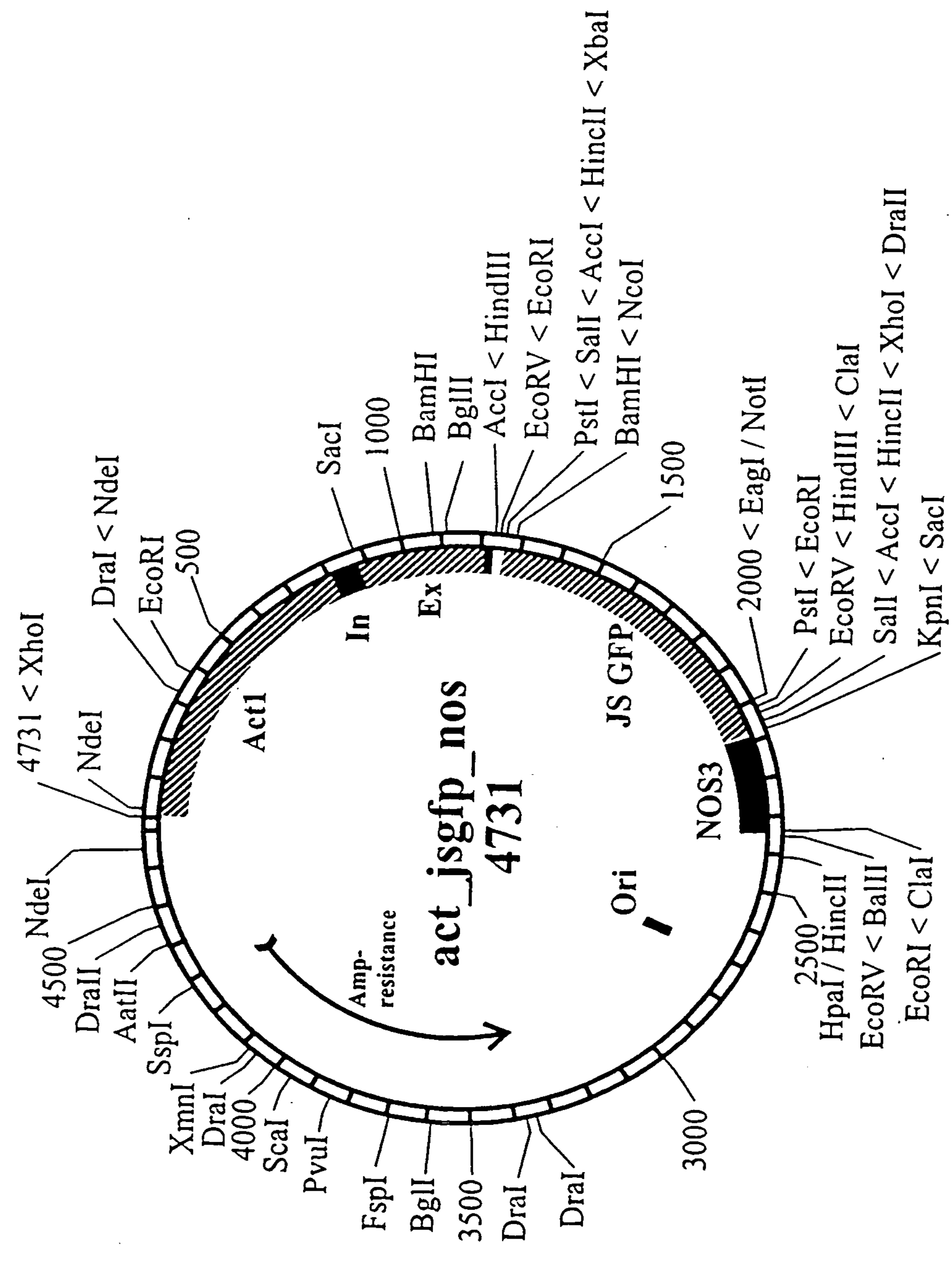
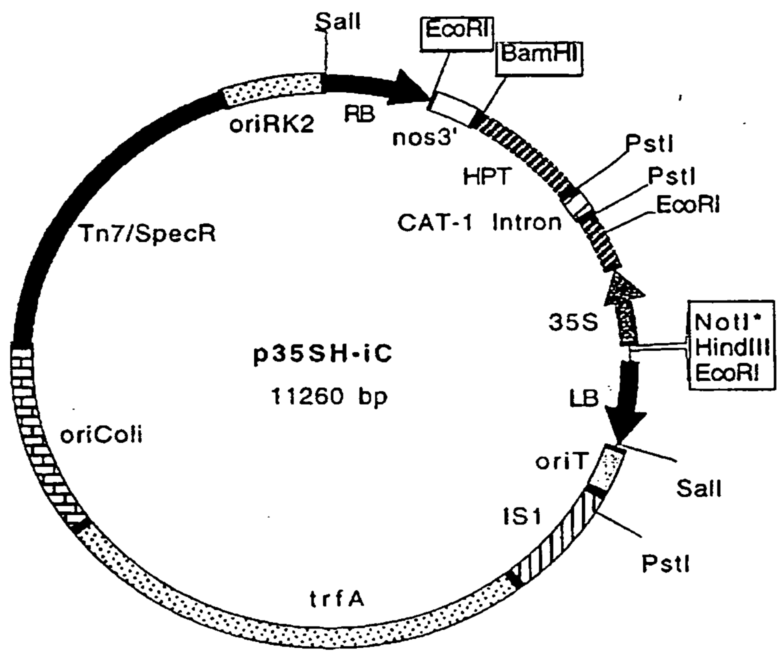

act_jsgfp_nos
4731
Act1
In
Ex
JS GFP
NOS3
Ori
Amp-
resistance
SacI
1000
BamHI
BglII
AccI < HindIII
EcoRV < EcoRI
PstI < SalI < AccI < HincII < XbaI
BamHI < NcoI
1500
2000 < EagI / NotI
PstI < EcoRI
EcoRV < HindIII < ClaI
SalI < AccI < HincII < XhoI < DraII
KpnI < SacI
2500
HpaI / HincII
EcoRV < BalII
EcoRI < ClaI
3000
DraI
DraI
3500
BglI
FspI
PvuI
ScaI
4000
DraI
XmnI
SspI
AatII
DraII
4500
NdeI
NdeI
4731 < XhoI
DraI < NdeI
EcoRI
500

Figure 22E


SalI
EcoRI
BamHI
oriRK2
RB
nos3'
HPT
PstI
PstI
EcoRI
CAT-1 Intron
Tn7/SpecR
35S
NotI*
HindIII
EcoRI
p35SH-iC
11260 bp
LB
oriColi
oriT
SalI
IS1
PstI
trfA

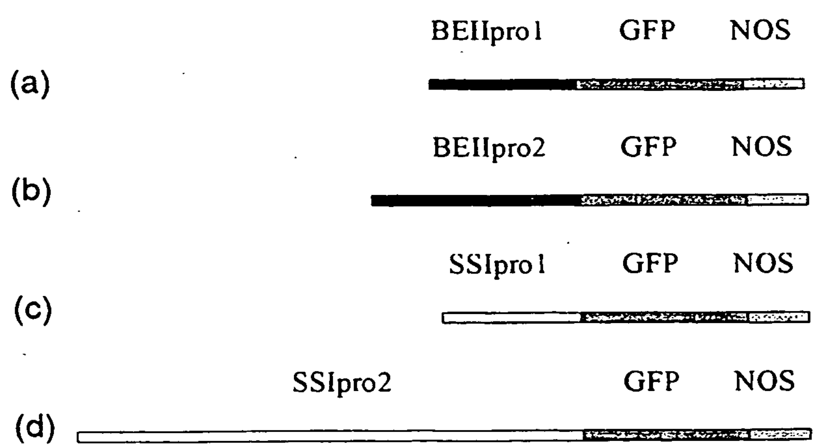

BEIIpro1 GFP NOS
(a)
BEIIpro2 GFP NOS
(b)
SSIpro1 GFP NOS
(c)
SSIpro2 GFP NOS
(d)

FIGURE 23

Primer Set	Key	Forward Primer	Forward Primer Sequence
1	E01'/E02	WBE2E1F	CGT CGC TGC TCC TCA GGA AG
2	E01/E02	sr854.1180F	CTG GCT GAC TCA ATC ACT ACG
3	E02/E03	WBE2E2F	CGC AAC CTG AAG AAT TAC AG
4	E03/E04	WBE2E3F	ATT TTC GGA GCC ATC TTG AC
5	E04/E05	WBE2E4F	TCG TGG TTA TGA AAA GCT TGG
6	E05/E06	sr913F	ATC ACT TAC CGA GAA TGG G
7	E05/I05	sr913F	ATC ACT TAC CGA GAA TGG G
8	E06/E07	WBE2E6F	ACA ATT GGA ATC CAA ATG CA
9	E07/E08	WBE2E7F	AGC TAT TCC TCA TGG CTC AC
10	E08/E09	WBE2E8F	TGC AGG CTC CAG GTG AAA TA
11	E10/E11	da5.seq	GGC TTG GAT ACA ATG CAG TGC
12	E12/E13	da151.seq	TTG ACG GCT TGA ATG GTT TC
13	E17/E18	WBE2E17F	TTT AGG TGG TGA AGG CTA TCT
14	E18/E19	sr860R	AAT GGA TAG ATT TTC CAA GAG G
15	E19_3'	WBE2-2395F	AGC AGA ACT GCG GTC GTG TA

Reverse Primer	Reverse Primer Sequence	Temp	bp
WBE2E2R	CAG GAC CTT CCC TGG AGA GG	57.4	401
WSBE9E2R	GGC ACG AGT GTG TGT ACC TGT A	57.7	601
sr866F	TAT CTT CAG GTA TCT ACA GC	49.8	309
WBE2E4R2	ATG CTT CCA ATC CAC CTT CA	-	>450
WBE2E5R	GAG CCC ATT CTC GGT AAG TGA	50.5	234
WBE2E6R	CTG CAT TTG GAT TCC AAT TG	49.9	232
WBE2I5R	CAG TAA GCT AGT TGG TGA ATA	46.6	106
WBE2E7R	GGG AGG AAA ATC TCC CAA AC	51.0	402
sr915F	CCA TTG AAA GGT ATT TCA CC	51.1	203
sr912F	TAA CTT ATT GAC ATA CCG G	48.4	439
WBE2E11R	CTG GAG TTC CAA AAC GGC TAC	51.2	289
WBE2E13R	ATT CTT CAA GCC ACC ATC TC	51.6	244
WBE2E18R	TAT TGT TAT TTC CAG GGG AGA	50.2	258
da23.seq	TGC TGC ATT GCC TGA TCG AA	50.4	~295
WBE2-2634R	AAC ACC CAG GCC CGT CCA TT	57.2	240

Figure 24

SBE II Intron 5 primer set - digested with Dde1

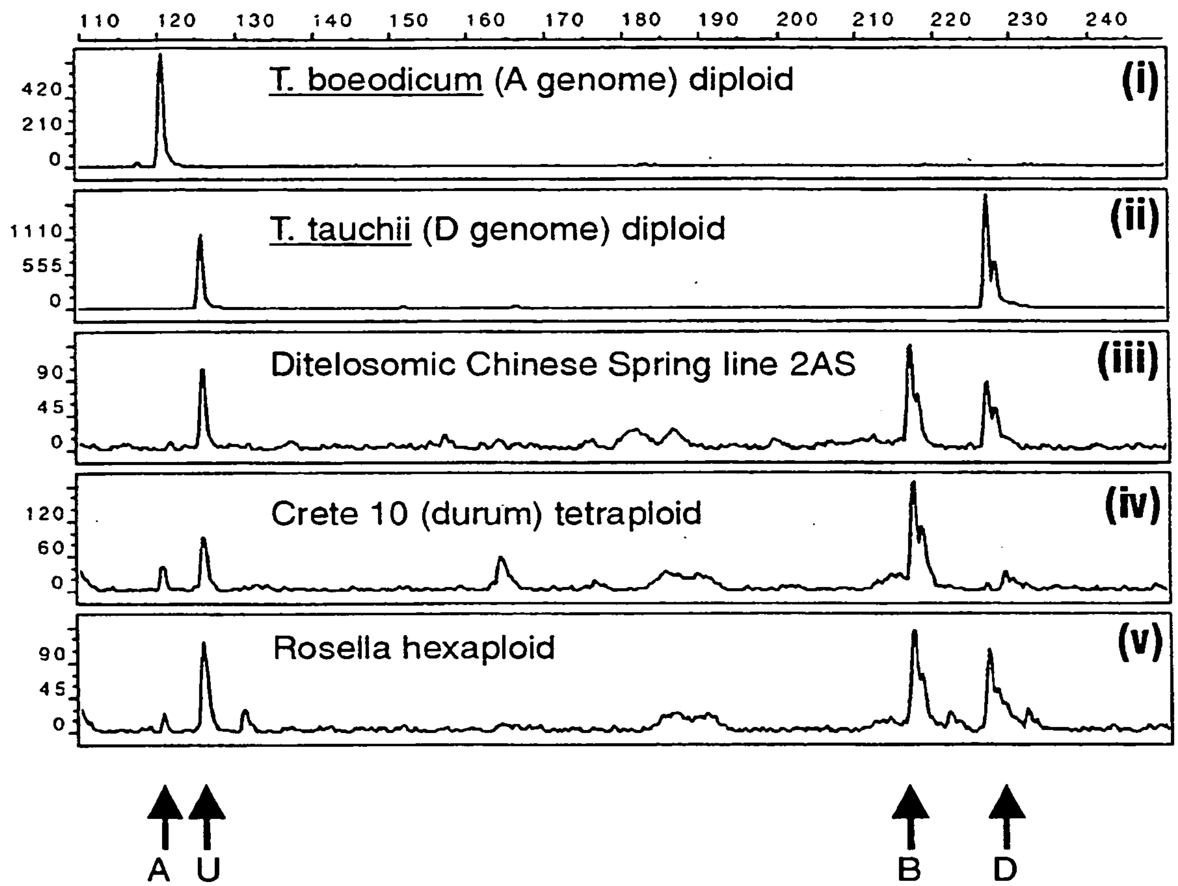


FIGURE 25

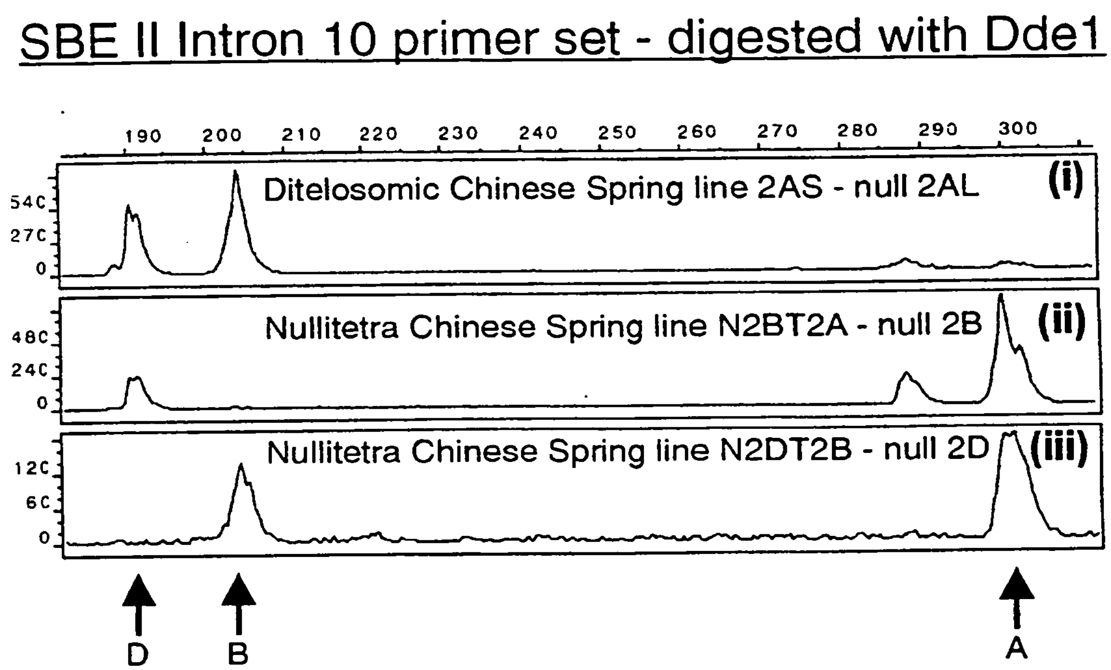

SBE II Intron 10 primer set - digested with Dde1
190
200
210
220
230
240
250
260
270
280
290
300
Ditelosomic Chinese Spring line 2AS - null 2AL
(i)
540
270
0
Nullitetra Chinese Spring line N2BT2A - null 2B
(ii)
480
240
0
Nullitetra Chinese Spring line N2DT2B - null 2D
(iii)
120
60
0
D
B
A

FIGURE 26

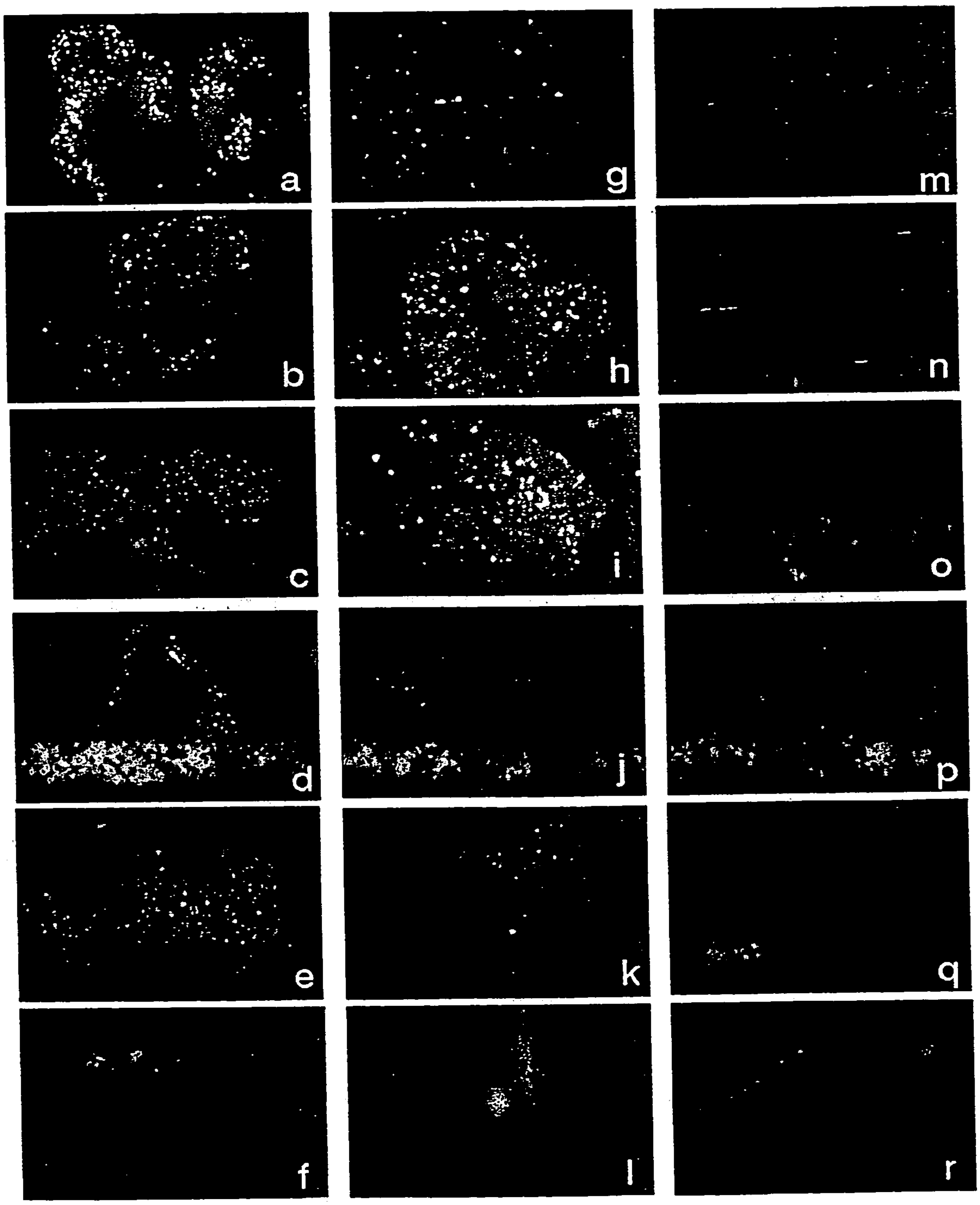


FIGURE 27

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU 98/00743

A.	CLASSIFICATION OF SUBJECT MATTER
Int Cl^{6}:	C12N 9/24, 15/55

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Electronic Data base box

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Electronic Data base box

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPAT - Starch branching enzyme #, promoter #, debranching enzyme : CA, medline - Starch Branching enzyme #, starch synthase, triticum, wheat: **Genebank, Embl** - sequences as claimed.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	AU-B-19028/95 (688006) (Nat. Starch &Chem) 17 October 1995. (See fig 8 in particular)	1, 2, 16, 21, 22 and 36
PX	AU-A 48747/97 (Nat. Starch & Chem) 14 May 1998. Epd 5 November 1996 (See Fig 4 in particular)	1, 2, 16, 21, & 22
X	WO 97/04113 (DANISCO A/S) 6 February 1997 (See fig 8 and page 22 in particular)	1, 2, 16, 21& 22

[X] Further documents are listed in the continuation of Box C

[X] See patent family annex

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 October 1998	20 OCT 1998

Name and mailing address of the ISA/AU	Authorized officer
AUSTRALIAN PATENT OFFICE PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No.: (02) 6285 3929	**PHILIPPA WYRDEMAN** Telephone No.: (02) 6283 2554

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU 98/00743

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	AU-B-65392/94 (693787) (DANISCO A/S) 8 November 1994. (See page 43 in particular)	1, 2, 16, 21 & 22
X	AU - A 77165/95 (AMYLOGENE HB) 5 June 1997 (See in particular seq. ID# 1, page 12)	1, 2, 16, 21 & 22
X	Nair, R. B et al (1997) PLANT SCIENCE "Isolation, characterisation and expression analysis of a starch branching enzyme II cDNA from wheat" vol. 122, pages 153-163. (See entire document)	1, 2, 16, 21, & 22

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/AU 98/00743

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member					
WO	9704113	AU	66146/96	EP	839203		
AU	94/65392	CA	2160159	EP	693128	GB	2291878
		NZ	265061	WO	9424292		
AU	95/77165	WO	97/20040	EP	863983	NO	982443
		SE	9601506	SE	9504272		
AU	95/19028	WO	9526407	EP	754235	CA	2186399
AU	97/48747	WO	9820145	GB	2320716		
GB	9307408	AU	65392/94	CA	2160159	EP	693128
		GB	2291878	NZ	265061	WO	9424292
SE	9504272	AU	77165/96	EP	863983	NO	982443
		SE	9601506	WO	9720040		
GB	9406022	AU	19028/95	CA	2186399	EP	754235
		WO	9526407				
GB	9623095	AU	48747/97	GB	2320716	WO	9820145

END OF ANNEX

THIS PAGE BLANK (USPTO)